

Essays in Development Economics and Econometrics

by
Robert J. Garlick

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Public Policy and Economics)
in the University of Michigan
2013

Doctoral Committee:

Professor David Lam, Co-chair
Professor Jeffrey A. Smith, Co-chair
Assistant Professor Manuela Angelucci
Professor John DiNardo
Professor Brian A. Jacob

© Robert J. Garlick 2013

All Rights Reserved

Acknowledgements

This dissertation would not have been started, let alone finished, without the assistance of many others. Starting from an early age, my parents Alison and Ken encouraged me to be curious about almost every topic imaginable and laid the seeds for an academic career. Anyone who saw the many family dinners spent in heated discussion and concluding with encyclopedia-reading could have predicted my career choice two decades ago. They were unstintingly supportive of my academic interests even when this took me far from home for the past six years. Given this upbringing, it's hardly surprising that my sister Julia is also well on her way to completing a doctorate in economics. She bore the brunt of my youthful argumentative tendencies but showed remarkable tolerance as I unintentionally prepared for the most combative seminar rooms. She has been an important part of my intellectual development and I look forward working with her in the future.

Many faculty members at the University of Cape Town nurtured my interest in economics, gave generously of their time, and helped me to access a wide range of research and teaching opportunities. I am very grateful to Justine Burns, Lawrence Edwards, Malcolm Keswell, Murray Leibbrandt, Sam Muradzikwa, Corne van Walbeek, Martin Wittenberg, and particularly Johann Fedderke. Johann, Martin, and Murray wrote the reference letters that were vital to my getting into graduate school.

My advisors at the University of Michigan – David Lam, Jeff Smith, Manuela Angelucci, John DiNardo and Brian Jacob – have been superb. Jeff and David read many of my papers, met with me regularly, and wrote more reference letters than I can count. Both have hugely influenced the way I think about economics and economic research. Any informed observer will see Jeff's influence on my methodological work and David's influence on my linkages between theory and data. They were also tremendously helpful during my “debutante ball”

into the economics profession over the past year. John was an endless source of encouragement, insight, and unorthodox but brilliant suggestions. He was incredibly generous with his time through a life-threatening illness and, although he will downplay his role at every opportunity, has been a vital contributor to my intellectual development. Brian supervised my first exposure to research, helped me to learn the less-than-glamorous process of data analysis, and always reminded me to keep it simple when I tried to overcomplicate things. Manuela kept me focused on the underlying economic models and emphasized the importance of building toward a coherent narrative.

Many other faculty members have provided valuable dissertation feedback over the past six years. John Bound in particular was a regular source of both academic and professional advice, for which I am very grateful. Raj Arunachalam, Martha Bailey, Charlie Brown, Matias Cattaneo, Sue Dynarski, Kevin Stange, Rebecca Thornton, and Dean Yang all provided useful input.

Mary Corcoran directed the dual economics and public policy program for most of my time at Michigan and was an exceptional and generous mentor. The graduate program administrators in economics and public policy – Mary Braun and Michelle Spornhauer – played an important role in allowing me to have a smooth and crisis-free ride through graduate school.

My fellow students were incredible sources of academic and personal support. I spent many hours discussing problem sets, research, and life with Emily Beam, Tanya Byker, Dave Cashin, Andrew Goodman-Bacon, Italo Gutierrez, Josh Hyman, Sasha Killewald, Elena Patel, and Nate Seegert. Elias Walsh, Jess Goldberg, and Todd Pugatch provided the valuable perspective of an older generation of graduate students.

Last, but very much not least, my girlfriend Jessi Streib provided years of companionship and support, incredible tolerance for my long hours in the office, extremely last-minute editing assistance, and regular reminders of the limits of economics. She made the last fifty-three months and fifteen days a much brighter time and I'm very glad we're one of the lucky academic couples who both found excellent jobs in the same place.

Contents

| | |
|---|-------------|
| Acknowledgements | ii |
| List of Tables | vi |
| List of Figures | viii |
| 1 Academic Peer Effects with Different Group Assignment Rules: Residential Tracking versus Random Assignment | 1 |
| 1.1 Introduction | 1 |
| 1.2 Experimental Setting and Research Design | 13 |
| 1.3 Treatment Effects of Tracking | 20 |
| 1.4 What Can Be Learned from Cross-Dormitory Variation? | 28 |
| 1.5 Mechanisms | 32 |
| 1.6 Conclusion | 35 |
| 1.A Robustness Checks | 53 |
| 1.B Reweighted Nonlinear Difference-in-differences Model | 60 |
| 2 How Price Sensitive is School Enrollment? Evidence from Nationwide School Fee Reforms in South Africa | 70 |
| 2.1 Introduction | 70 |
| 2.2 Theoretical Framework | 76 |
| 2.3 Background Information and Identification Strategy | 79 |
| 2.4 Effect of fee elimination on enrollment | 88 |
| 2.5 Explaining the price-insensitive demand | 93 |

| | | |
|----------|--|------------|
| 2.6 | Effects of fee elimination on school composition and outcomes | 97 |
| 2.7 | Conclusion | 99 |
| 3 | Mobility Treatment Effects: Identification, Estimation, and Application | 113 |
| 3.1 | Introduction | 113 |
| 3.2 | Identification | 117 |
| 3.3 | Estimation | 126 |
| 3.4 | Application | 127 |
| 3.5 | Conclusion | 130 |
| | Bibliography | 132 |

List of Tables

| | | |
|-----|--|----|
| 1.1 | Baseline demographics and high school graduation test scores for dormitory and non-dormitory students in the tracking and random assignment periods | 38 |
| 1.2 | Estimates of the average treatment effect of tracking using a linear difference-in-differences regression model | 39 |
| 1.3 | Estimates of the average treatment effect of tracking for students in each quartile of the high school graduation test score distribution using a linear difference-in-differences regression model | 39 |
| 1.4 | Estimates of the average treatment effect of tracking for race and gender subgroups using a linear difference-in-differences regression model | 40 |
| 1.5 | Estimates of the average and inequality treatment effects of tracking using a nonlinear difference-in-differences model with and without reweighting to adjust for differences in baseline student characteristics. | 40 |
| 1.6 | Transition probabilities for students from their rank in the distribution of high school graduation test scores to their rank in the distribution of university GPA | 41 |
| 1.7 | Treatment effects of tracking on students' academic mobility, measured by the probability of changing their rank in the distribution of high school graduation test scores to their rank in the distribution of university GPA | 41 |
| 1.8 | Estimates of the effect of peers' high school graduation test scores on students' university GPAs using a linear-in-means model for peer effects | 42 |
| 1.9 | Estimates of the effect of peers' high school graduation test scores on students' university GPAs using a quadratic-in-means model for peer effects | 42 |

| | | |
|------|---|-----|
| 1.10 | Estimates of the effect of peers' high school graduation test scores on students' university GPAs, separately for peers in the students' own and different race groups | 43 |
| 1.11 | Estimates of the effect of peers' high school graduation test scores on students' university GPAs, separately for peers in the students' own and different college/faculty/school (e.g. commerce, engineering, social science) | 44 |
| 1.12 | Average treatment effects of tracking on different measures of university academic performance estimated using a linear difference-in-differences regression model | 45 |
| 1.13 | Estimates of the average treatment effect of tracking within each faculty/school/college using a linear difference-in-differences regression model | 66 |
| 1.14 | Estimates of the average treatment effect of tracking within the six largest entry-level classes using a linear difference-in-differences regression model . . | 67 |
| 1.15 | Instrumental variables estimates of the average treatment effect of tracking using a linear difference-in-differences model, instrumenting the year of college admission with the year of high school graduation and dormitory/non-dormitory status with an indicator for whether the student attended a high school in Cape Town | 67 |
| 1.16 | Placebo difference-in-differences tests for differences in the pre-treatment trends for dormitory and non-dormitory students | 68 |
| 1.17 | Estimated p -values for the test that the average treatment effect of tracking is zero using alternative test procedures in a linear difference-in-differences regression model | 68 |
| 2.1 | Treatment effects of fee elimination on school-level enrollment | 101 |
| 2.2 | Treatment effects of fee elimination on enrollment in high- and very-high poverty schools (quintiles 2 and 1) | 102 |
| 2.3 | Treatment effects of fee elimination on enrollment in rural and urban schools | 103 |
| 2.4 | Treatment effects of fee elimination on measures of school resources, grade progress, and composition | 104 |

List of Figures

| | | |
|-----|--|----|
| 1.1 | Difference between tracking and randomization periods in the mean high school graduation test scores of students' dormitory peer groups for students at each percentile of the high school graduation test score distribution. . . . | 46 |
| 1.2 | Mean and standard deviation of high school graduation test scores in each dormitory-by-year | 47 |
| 1.3 | Cumulative distribution function of high school graduation test scores for dormitory and non-dormitory students in the tracking and random assignment periods | 48 |
| 1.4 | Average treatment effects of tracking for students at each percentile of the distribution of high school graduation test scores | 49 |
| 1.5 | Nonlinear difference-in-differences analysis, comparing the observed distribution of GPA for tracked dormitory students to the counterfactual distribution that the same students would have experienced if they had not been tracked | 50 |
| 1.6 | Nonlinear difference-in-differences analysis, which uses the observed and counterfactual distributions of GPA for tracked dormitory students to estimate quantile treatment effects of tracking | 51 |
| 1.7 | Density of selected summary statistics for high school graduation test scores at the dormitory-by-year level, showing how these densities differ between the tracking and random assignment periods | 52 |
| 1.8 | "Bias-corrected" estimates of the average treatment effect of tracking from a linear difference-in-differences regression model with different assumptions about the relationship between an unobserved scalar and GPA and the time trend in the unobserved variable | 69 |

| | | |
|-----|---|-----|
| 2.1 | Conceptual framework showing the effect of fee elimination $C_1 \rightarrow C_2$ on the proportion of students enrolled P | 105 |
| 2.2 | Falsification test for a discontinuity in the density of standardized poverty scores | 106 |
| 2.3 | Time trend in net enrollment rate | 107 |
| 2.4 | Probability that schools eliminate fees, by poverty score | 108 |
| 2.5 | Change in enrollment from 2005/6 to 2007/8, by poverty score | 109 |
| 2.6 | Level of enrollment from 2003 to 2008 for intention to treat schools, control schools, and reweighted control schools | 110 |
| 2.7 | Treatment effects of fee elimination on enrollment by grade | 111 |
| 2.8 | Changes in enrollment from 2005 to 2006 and 2006 to 2007 for fee-charging schools located at different distances from fee-eliminating schools | 112 |

Chapter 1

Academic Peer Effects with Different Group Assignment Rules: Residential Tracking versus Random Assignment

1.1 Introduction

Group structures are ubiquitous in education and group composition may have important effects on education outcomes. Students in different classrooms, living environments, schools, and social groups are exposed to different peer groups, receive different education inputs, and face differential institutional environments. A growing body of empirical evidence shows that students' peer groups influence their education outcomes even when resource and institutional differences across groups are negligible. Peer effects have been documented on students' college GPAs (Sacerdote, 2001), standardized test scores (Hoxby, 2000), college entrance examinations (Ding and Lehrer, 2007), study habits (Stinebrickner and Stinebrickner, 2006), major choices (Di Giorgi, Pellizzari, and Redaelli, 2010), job search (Marmaros and Sacerdote, 2002), and social networks (Marmaros and Sacerdote, 2006). Academic peer effects play a role in both empirical and theoretical studies of classroom tracking (Arnott, 1987; Duflo, Dupas, and Kremer, 2011), neighborhood segregation (Benabou, 1996; Kling, Liebman, and Katz, 2007), school choice and vouchers (Epple and Romano, 1998; Hsieh and Urquiola, 2006), and school integration and busing policies (Angrist and Lang, 2004).¹ The majority of these studies focus on the effect of assignment to or selection into different peer

¹These estimates may be sensitive to measurement of peer characteristics or peer groups. Administrative data used to define peer groups may imperfectly measure the set of peers with whom students actually interact, leading to biased estimates of peer effects (Foster, 2006). Different measures of peer characteristics may also lead to different conclusions about the strength of peer effects: Stinebrickner and Stinebrickner (2006) estimate much larger effects on college students' GPAs from their roommates' high school GPA than SAT scores.

groups for a given group assignment or selection process.

This paper asks a subtly different question: What are the effects of different policies for assigning students to groups? This contributes to a small but growing empirical literature on optimal group design. Comparison of different group assignment policies corresponds to a clear social planning problem: How should students be assigned to groups in order to maximize some measure of academic output, subject to a given distribution of student characteristics? Different group assignment policies leave the marginal distribution of inputs into the education production process unchanged. This raises the possibility of increasing academic output without the pecuniary costs associated with most education interventions. Such low cost education interventions are particularly attractive for developing country education systems that often face serious resource shortages. I study the relative effects of two common policies for assigning college students to residential groups: random assignment and academic tracking. Under random assignment, each residential group is approximately representative of the entire population of students. Under tracking, each student is placed in a residential group of peers with similar baseline academic proficiency. I use a natural experiment in the dormitory system of the University of Cape Town in South Africa to show that tracking substantially harms low-scoring students and has little effect on high-scoring students. The net effect is a decrease in mean academic output and an increase in inequality of academic output.

This cross-policy comparison is important because peer effects estimated under one assignment policy may not correctly predict the effects of a change in the assignment policy. Such prediction requires strong assumptions about out-of-sample extrapolation, model specification, and economic behavior. For example, randomly assigning students to dormitories allows researchers to identify the effect of residential peers' academic proficiency on students' academic outcomes. These "within-policy" estimates can predict how a student's outcome will change if the proportion of high ability students in her dormitory increases marginally, for example from 40% to 60%. However, there is a low probability that random assignment will generate "tracking-like" dormitories where the proportion of high ability students is near 0 or 100%.² Any attempt to predict the effect of tracking students into dormitories

²In contrast, random assignment to smaller groups such as roommate pairs may plausibly generate sub-

thus relies on out-of-sample extrapolation. This problem is exacerbated by the fact that model misspecification is a real danger in peer effects estimation. Students' outcomes may depend on many different features of the distribution of peers' characteristics (mean, variance, order statistics, etc.) and models estimated under one assignment policy may lead researchers to incorrect conclusions about which features of this distribution matter under another assignment policy.

Even if the model is statistically correctly specified, there may be economic problems with predicting the effects of changes in group assignment policies. Residential peer effects may operate through two conceptually distinct channels: a direct channel, in which peers exert influence through spatial proximity,³ and an indirect channel, in which spatial proximity influences the formation of social and/or academic networks, which in turn influence student outcomes. Most studies of peer effects under random assignment estimate a reduced form combination of these effects. However, changes in dormitory assignment policies might change the relationship between spatial proximity and social networks, in which case the indirect effect will not be policy-invariant. For example, students' propensity to form intra-dormitory social ties might be higher under tracking than random assignment, which magnifies the indirect effect under tracking relative to random assignment. Estimates that combine direct and indirect effects would then depend on assignment policies and so could not reliably predict the effect of changes in these policies.

The empirical and methodological literature on this topic is sparse. Bhattacharya (2009) and Graham, Imbens, and Ridder (2011) develop methods for predicting the effect of various assignment policies using peer effects estimated under conditionally random assignment. However, these approaches impose strong assumptions on the data generating process. For example, Graham, Imbens, and Ridder (2011) limit their attention to cases where peer types are binary, so the distribution of peer characteristics in a group is fully characterised by the mean. Their model only applies to cases where out-of-sample extrapolation is not

stantial numbers of "high-high" and "low-low" pairs. These pairs could in principle be used to predict the effect of changing the assignment policy to tracking, as in Bhattacharya (2009). I thank Todd Stinebrickner for pointing out this distinction. The concern about policy-invariance still applies in this smaller group setting.

³For example, noisy neighbors in the dormitory might impair students' ability to study, even if no direct interaction occurs between the students.

required and they effectively assume that the policy sensitivity problem discussed above does not apply. These papers still represent substantial methodological contributions – the strong assumptions that they make merely reflect the difficulty of the cross-policy prediction problem. Carrell, Sacerdote, and West (2012) and Duflo, Dupas, and Kremer (2011) provide what appears to be the only direct empirical evidence regarding the effect of different policies for creating peer groups. Both studies find that peer effects estimated under one assignment policy do not predict the result of changing that assignment policy. The key advantage of my study, and the work by Carrell, Sacerdote, and West (2012) and Duflo, Dupas, and Kremer (2011), is that empirical cross-policy comparison avoids the need to estimate policy-invariant and extrapolatable peer effects. Instead, I can estimate the relative effects of different policies and then use these results to learn about the structure of the underlying peer effects.

A better understanding of how peer effects operate under different group assignment policies may shed light on important policy debates in education. In particular, an extensive empirical and theoretical literature has studied the effect of tracking students into classrooms or schools (Betts, 2011). Most empirical studies estimate reduced form effects of tracking relative to other assignment policies but cannot separate the mechanisms driving their results. The results may be driven by peer effects or by differences between tracked and untracked groups in curricula, instruction, or other resources. A small number of papers use careful data collection strategies to assess the relative importance of different mechanisms (Ding and Lehrer, 2007; Duflo, Dupas, and Kremer, 2011; Pop-Eleches and Urquiola, 2012) but the role of peer effects in academic tracking remains poorly understood. Group assignment or selection processes also play potentially important roles in debates around school choice (Epple and Romano, 1998; Hsieh and Urquiola, 2006) and school integration (Angrist and Lang, 2004).

This paper provides a bridge between the literatures on academic tracking and on optimal design of peer groups. I study a natural experiment where students were assigned to residential groups (dormitories) using tracking for several years and thereafter by random assignment. I contrast the distribution of dormitory students' academic outcomes under the two policies and show that mean academic performance was lower and inequality higher under tracking. I argue that baseline differences between the tracked and untracked students

are small and show that the results are highly robust to accounting for these differences. Specifically, I use non-dormitory students who live in private accommodation as a control group to remove any time trends, flexibly control for differences in students’ baseline observed characteristics, and use university admission policies to construct instrumental variables to control for differences in students’ unobserved baseline characteristics. The results are highly robust to these strategies and to several additional sensitivity analyses, including a placebo test using data from before the policy change. My findings show that alternative peer group assignment policies can have substantial effects on academic performance. In particular, the distribution of outcomes under tracking is worse than under random assignment for most plausible social welfare criteria in this setting. The tracked and untracked groups do not experience different curricula, instruction or institutional environments. While there are differences in physical facilities across dormitories, these are stable through time and so do not differ between the tracked and untracked groups. The differences in academic performance must thus be driven by peer effects, suggesting that a “pure peer effects” mechanism is an important part of the effect of tracking in other settings.

I also explore the relationship between peer effects implied by the cross-policy results and peer effects estimated under a given policy. In particular, random student assignment to dormitories allows me to consistently estimate the effect of peers’ baseline characteristics on students’ academic outcomes. These estimates confirm that students are affected by their residential peers’ academic proficiency, measured by high school graduation test scores, and that low-scoring students are substantially more affected than high-scoring students. This pattern is qualitatively consistent with the results of the main cross-policy analysis but predicts far smaller effects of tracking than I observe. This may reflect out-of-sample prediction problems or behavioral responses by students. I show evidence consistent with the former explanation by documenting that the effect of peer characteristics on student outcomes is much stronger when those peers are of the student’s own race group. This suggests that spatial proximity generates peer effects but that the relationship is not mechanical. Instead, students are influenced more by peers living near them if those peers are also socially proximate (crudely proxied by race, which is a salient feature of social networks in South Africa). This suggests that the indirect channel (from spatial proximity to interaction to academic

outcomes) is an important part of peer effects in this setting. Hence, peer effects estimated under random assignment might fail to predict the effect of tracking because they cannot capture changes in the strength of this indirect effect.

This paper’s substantive contribution is facilitated by several advanced econometric techniques. Standard methods for comparing academic outcomes under two different policies typically focus on the average treatment effect, for all students or for particular subgroups of interest. This is informative but more can be learned by contrasting the full distributions of outcomes. I therefore estimate the counterfactual distribution of academic outcomes that tracked students would have achieved if tracking were not implemented, using the “nonlinear difference-in-differences” model proposed by Athey and Imbens (2006). I also extend the model to flexibly account for differences in the distribution of observed student characteristics such as race and gender. Having constructed the full counterfactual distribution of outcomes, I can estimate the full set of quantile treatment effects of tracking. I can also calculate summary statistics of interest on the observed and counterfactual distribution and so estimate the effect of tracking on various measures of academic inequality. I also note that tracking may affect students’ academic mobility, measured by the probability that low-ranked students in high school will become high-ranked in university and vice versa. I develop the idea of a “mobility treatment effect” to quantify this effect and develop the appropriate statistical theory in a companion paper (Garlick, 2012).

These findings suggest that group assignment policies may have substantial effects on academic performance even when other education inputs are held constant. Such policy changes impose few pecuniary costs and so offer policymakers an attractive method for boosting academic performance. However, Pareto ranking of different policies may not be possible, as different group assignment policies impose transfers across different groups of students and academic outputs are not tradeable. In this setting, the large losses that tracking imposes on low-scoring students are not accompanied by significant positive effects on high-scoring students. Most social welfare criteria would thus prefer random assignment to tracking but care should be taken in applying this result in other settings. For example, the distribution of students’ academic proficiency at the University of Cape Town is wider than in most tertiary institutions in developed countries. The university attracts many of the

highest performing students in South Africa but also admits many academically promising students from low-performing high schools with weak preparation for further study.⁴ These are the students most affected by tracking, so my results are particularly relevant to policy debates about how well academically underprepared students perform in tertiary education (Bertrand, Hanna, and Mullainathan, 2010; Frisanco and Krishna, 2011).

Organization of the paper: Section 1.2 discusses the setting that I study and the two assignment policies. Students were assigned to dormitories using a tracking policy up to the 2005 academic year and randomly assigned to dormitories from the 2006 academic year onward. I therefore employ a difference-in-differences design that compares the academic outcomes of dormitory students under the two assignment policies, using non-dormitory students to control for time trends in students' academic outcomes. I briefly discuss some potential threats to the validity of this design and show in the appendices that my results are highly robust to such concerns. This section also quantifies the change in the distribution of peer characteristics across groups, showing that the policy induced a large reallocation in peer quality.

Section 1.3, which compares students' outcomes across the tracking and random assignment policies, is the main empirical component of the paper. I find that tracking reduces mean GPA by approximately 0.12 standard deviations. This result is driven by a large negative effect on the students with low high school graduation test scores (up to 0.25 standard deviations) and a small and imprecisely estimated positive effect on the upper tail. Quantile treatment effects verify that tracking substantially lowers the left tail of the GPA distribution and has almost no effect on the right tail. In fact, the GPA distribution under tracking is almost stochastically dominated by the counterfactual GPA distribution that would have arisen in the absence of tracking. I show that this pattern of treatment effects implies substantial increases in standard measures of inequality, suggesting that tracking has negative efficiency and equity effects. However, I find no effect of tracking on academic mobility, measured by the probability that students change their rank in the distribution of academic achievement from high school to university. This suggests that tracking disproportionately

⁴These students perform poorly in the standardized high school graduation tests that I use to measure academic proficiency. This may reflect a combination of poor academic preparation and low ability, rather than any one particular factor.

reduces low scoring students' level of GPA but leaves their rank largely unchanged.

Section 1.4 explores what could have been learned about peer effects without a policy change, if I had only observed students during the random assignment period. I exploit the fact that random assignment to dormitories balances peer group means only in expectation and estimate the effect of random variation in dormitory-level measures of student characteristics. I show that students' GPAs are affected by the characteristics of their residential peers and that students with low high school graduation test scores are more strongly affected. I use empirical and theoretical arguments to highlight the difficulty of predicting the effects of tracking using these results.

Section 1.5 explores the economic behavior driving the negative effect of tracking. I begin by showing that peer effects operate mainly within race groups. I interpret this as evidence that dormitory assignment influences student outcomes indirectly. Spatial proximity influences students' interaction patterns but spatially proximate peers are more influential if they are also socially proximate (crudely proxied by being members of the same race group). The strength of this indirect peer effects channel may depend on the group assignment policy, which reinforces the difficulty of cross-policy prediction. I then go on to show that peer effects are not stronger between students taking the same courses, which I interpret as evidence that the relevant type of student interaction is not direct academic collaboration (completing problem sets together, copying essays, etc.). Finally, I use an unusual feature of the timing of assessment at the University of Cape Town to show that peer effects operate from early in the semester, not just during exam periods.

Section 1.6 concludes and discusses the implications of these results. I argue that this paper provides important and novel evidence on the academic effects of residential tracking, which are driven entirely by differences in peer groups across dormitories. This finding is most directly relevant to policymakers who assign students to residential groups. It also indicates that a pure peer effect mechanism may be important in other group assignment problems. For example, existing studies on academic tracking by classroom or school typically conflate differences in peer group characteristics with differences in instructor behavior or school resources. My results provide direct evidence regarding the importance of group composition in evaluating the desirability of tracking policies. This result is also relevant to the literature

on neighborhood segregation by family socioeconomic status (typically highly correlated with academic proficiency measures used for tracking).

The appendices present a variety of robustness checks to show that the results in section 1.3 are not driven by violations of the identifying assumption underlying the difference-in-differences design. This design assumes that the time trend in students' academic outcomes between the tracking and randomization periods would have been equal for dormitory students (the treatment group) and non-dormitory students (the control group) if the policy change had not occurred. Appendix 1.A.1 shows that the results are not driven by changes in course-taking behavior between the two periods. Appendix 1.A.2 shows that the results are not driven by students strategically selecting whether or not to live in the dormitory system. I construct an instrument for treatment status using the university's admissions policies and show that the instrumental variables and least squares estimates are very similar. Appendix 1.A.3 shows that the trends in student outcomes and baseline characteristics were similar for dormitory and non-dormitory students before the policy change. This placebo test provides reassuring evidence that the results are not driven by different time trends in the treatment and control groups. Appendix 1.A.4 presents a formal sensitivity analysis showing that an implausibly large change in students' unobserved characteristics would be needed to explain the observed treatment effects. Finally, appendix 1.A.5 assumes that the treatment effects have been correctly estimated and shows that inferences regarding these effects are robust to alternative test procedures.

Appendix 1.B provides more detail regarding the nonlinear difference-in-differences model employed in section 1.3. I present an intuitive explanation of the model's identification results and discuss how the identifying assumptions can be relaxed to accommodate differences in students' baseline observed characteristics.

Related literature: This paper relates to two literatures in the economics of education: peer effects and academic tracking. I study the effect of tracking students into groups whose only substantial difference is their composition. Unlike most studies of tracking, groups do not face different instructors, curricula or resources. I therefore estimate a treatment effect of tracking arising directly and exclusively from a shift in peer group composition and so provide a bridge between the tracking literature and the recent peer effects literature on

optimal group assignment. This section highlights some particularly relevant results from these literatures. For more comprehensive reviews of the literature, see Epple and Romano (2011) and Sacerdote (2011) on peer effects in education and Betts (2011) on academic tracking.

Manski (1993) provides the first formal treatment of the methodological challenges involved in estimating peer effects. He notes that in a linear regression of own academic outcomes on peers' baseline characteristics, a non-zero coefficient may arise for multiple reasons. The coefficient may reflect a causal effect of peer characteristics on own outcomes but may also reflect correlated unobserved characteristics due to endogenous group formation or correlated shocks because students are exposed to the same environmental and institutional factors as their peers.⁵ Most empirical papers therefore study settings in which students' peer groups vary for reasons uncorrelated with their own unobserved characteristics and include higher-level fixed effects to control for correlated shocks. The most directly relevant set of papers study the effect of random assignment to roommates with different SAT scores while using dormitory-level fixed effects to address correlated shocks (Sacerdote, 2001; Stinebrickner and Stinebrickner, 2006; Zimmerman, 2003).⁶ These studies typically find small to moderate effects of peer characteristics on academic outcomes. Cooley (2010) and Stinebrickner and Stinebrickner (2006) emphasize that estimates of peer effects may be sensitive to how peer characteristics are measured. Hoxby and Weingarth (2006) and several more recent papers argue that the functional form of reduced form estimating equations have substantial implications for interpretation of the results. Foster (2006) and Carrell, Fullerton, and West (2009) note that administrative units such as dormitories or classrooms may be poor proxies for students' true peer groups, which are typically unobserved by researchers.⁷ I discuss the relevance of these considerations to this paper in sections 1.4 and

⁵Manski also draws a distinction between two types of causal peer effects: exogenous effects, in which students are influenced by their peers' baseline characteristics, and endogenous peer effects, in which students' are influenced by their peers' outcomes. I return to this distinction in section 1.4.

⁶Other studies estimate the effect of cohort-level variation within school-by-grade units in gender composition, race composition or test scores (Hanushek, Kain, Markman, and Rivkin, 2008; Hoxby, 2000; Lavy and Schlosser, 2011) or the effect of approximately random assignment of students to classrooms (Ammermueller and Pischke, 2009; Vigdor and Nechyba, 2007; Kang, 2007). Another literature uses natural experiments in which students' peer groups are changed by school integration policies (Angrist and Lang, 2004; Hoxby and Weingarth, 2006) or natural disasters (Imberman, Kugler, and Sacerdote, 2012).

⁷Guryan, Jacob, Klopfer, and Groff (2008) is one the few papers that observes student interactions directly

1.5.

This paper is closely related to Carrell, Sacerdote, and West (2012), which examines the relative effects of two different policies for assigning freshmen to squadrons at the US Air Force Academy. They find that randomly assigning students leads to higher average GPAs than “mixing” students, so that some groups comprise only high and low ability students, while others comprise only students of moderate ability. However, using peer effects estimated across randomly created groups predicted that GPAs would be higher under the latter policy than the former. This result highlights the danger of predicting the effects of one policy using results estimated under another policy and underscores the value of papers that explicitly compare outcomes across different policies.

The tracking literature can be divided into two broad strands: the first studies the effect of tracking compared to other group assignment policies and the second studies the effect of assignment to different tracks. Betts (2011) reviews the first literature, noting that tracked and untracked groups often differ in curriculum, instructor characteristics, and school resources. Few papers are able to disentangle the relative importance of these factors and so cannot determine whether differences in the distribution of peer characteristics across tracked and untracked groups lead to differences in outcomes. Duflo, Dupas, and Kremer (2011) provide an important exception: they conduct a field experiment in which Kenyan primary schools are randomly assigned to two groups, one of which tracks first graders into classrooms and one of which randomly assigns them into classrooms. They find that tracking increases test scores for both high- and low-track students relative to random assignment. They also estimate that within the random assignment classrooms, assignment to a classroom with marginally higher scoring peers raises students’ test scores. They reconcile these results by arguing that students in the low track classrooms gained more from focused instruction than they lost from weaker peer groups.

Papers in the second strand of the tracking literature typically estimate the effect of assignment to a high instead of low track. Tracks often differ along multiple dimensions – including peer characteristics, curriculum, and instructor characteristics – so these designs do not directly test whether peers matter for academic outcomes. Some important exceptions instead of proxying them with group assignment.

include Ding and Lehrer (2007) and Pop-Eleches and Urquiola (2012) who study selective high schools in China and Romania respectively. The former paper finds large effects of attending selective schools and attributes much of this to measured teacher quality and a small portion to peers. The latter paper finds large effects of attending selective schools and selective tracks within schools but notes that qualification for higher tracks is associated with substantial changes in instructor, parent, and student behavior measured by detailed survey data.

This paper’s primary contribution is to the empirical literature but it also relates to a rich theoretical literature relating to peer effects and tracking. The theoretical discussion in section 1.4 builds off models of peer interaction discussed by Benabou (1996) and Epple and Romano (2011). Blume, Brock, Durlauf, and Ioannides (2011) and Brock and Durlauf (2011) review the literature on social interactions models. They place particular emphasis on models that attach a theoretical interpretation to the common estimands in the empirical literature. Graham (2011) reviews the literature on econometric methods for identification and estimation of such models.

The paper’s empirical methodology builds off a number of recent results in the heterogeneous treatment effects literature. I estimate quantile treatment effects of tracking using a nonlinear difference-in-differences model that recovers the full counterfactual distribution of GPAs that the tracked students would have obtained if they had instead been assigned to dormitories using random assignment (Athey and Imbens, 2006). I also implement an extension to the original Athey-Imbens models that incorporates changes in the distribution of observed student characteristics and discuss this in detail in appendix 1.B. This extension closely follows the reweighting methods proposed by DiNardo, Fortin, and Lemieux (1996) and formalized by Hahn (1998) and Firpo (2007), amongst others. I use this counterfactual distribution of GPAs to estimate the effect of tracking on inequality, following Firpo (2010). The treatment effects of tracking on academic mobility that I present in section 1.3 are based on identification and estimation results developed in a companion paper (Garlick, 2012).

1.2 Experimental Setting and Research Design

Experimental setting: The policy experiment I study was carried out at the University of Cape Town (UCT) in South Africa. UCT is a selective research university whose student body is not representative of South Africa’s population. However, the student body is substantially more heterogeneous than in US universities in which much of the prior peer effects literature is located. The university admits between 3500 and 4000 incoming students each year. Approximately half of these students live in private accommodation and half live in the 16 university dormitories, each of which accommodates between 35 and 237 students.⁸ The dormitories are similar to those found in many residential universities in the United States – they provide students with accommodation, meals, organized social activities, and some formal mentorship programs. Classes are not organized on dormitory lines, so students from different dormitories attend the same lectures. Incoming students before the 2006 academic year were tracked into dormitories based on their performance in a standardized high school graduation exam, with high-scoring students living in different dormitories to low-scoring students. From 2006 onward, students were assigned to dormitories by a random number generator. Under both policies, dormitory assignment did not directly affect students’ courses or instruction. Students from all dormitories and non-dormitory students were taught the same material in the same classes by the same instructors, so the dormitory assignment effect operated solely through students’ living environments and spatial peer groups. There is some variation in dormitories’ physical facilities and proximity to campus and to leisure opportunities but controlling for these differences using dormitory fixed effects makes little difference to the results.

The policy change: Under the tracking policy, students were assigned to dormitories based on their result in a standardized high school graduation examination, which was also used for admissions decisions.⁹ There were, however, two deviations from a pure tracking

⁸The mean dormitory size over the study period is 123 students and the 10th, 50th and 90th percentiles are 50, 112 and 216 students. There are no substantial changes in dormitory size between the two periods. Fourteen dormitories were open under both policies, one was open only under the tracking policy and one was open only under the random assignment policy. My results are robust to using only the fourteen dormitories open under both policies.

⁹The university’s admissions office converts international students’ results in A-level, International Baccalaureate (IB), and Higher International General Certificate in Secondary Education (HIGCSE) examina-

assignment, in which dormitories partition the distribution of high school graduation test scores. First, the assignment policy included an affirmative action component, so black students required lower scores for assignment to high track dormitories than white students.¹⁰ Second, students who applied late to the university were waitlisted for places in the dormitories and assigned when admitted students declined their admission offer or deregistered. The student at the top of the waitlist was assigned to the first space that became available, without regard to their high school graduation test scores. These exceptions increase within-dormitory and decrease between-dormitory variation in baseline test scores relative to pure tracking. This may attenuate the estimated treatment effect of tracking. Furthermore, the race-specific thresholds and *ad hoc* assignment of late applicants mean that there is non-trivial overlap in student’s high school graduation test scores across dormitories. There is thus insufficient mass near the dormitory-specific thresholds to allow me to estimate the effect of assignment to one dormitory or another by a regression discontinuity design.

Under the random assignment policy, students were assigned to dormitories using a race-blind random number generator. The staff member responsible for assignments regularly checked the racial composition of each dormitory and if he believed that one race group was underrepresented, he manually assigned the next few applicants of the underrepresented race to that dormitory. The criteria for determining “underrepresentation” were left to the discretion of the staff member, who reported making “only a few” reassignments each year.

This raises the possible concern that students were able to manipulate their dormitory assignment by lobbying staff members involved in the assignment process. When interviewed, the director of the admissions office acknowledged that this was a risk under both the tracking and random assignment regimes but stated that “everyone involved in the process was instructed to present a united front ... that residence assignments were final.” Informal discussion with both students and staff suggest that this policy was strictly enforced. Such manipulation, if it took place, complicates interpretation of the empirical results but does not affect the internal validity of my research design.

tions into equivalent results on the South African graduation examination.

¹⁰South Africa’s population was divided into four groups under *apartheid*’s racial classification system: black, coloured, Indian, and white. Given the ongoing salience of racial divisions, these distinctions are still in widespread use in social science research, public discourse, and government policy.

Under both the tracking and random assignment policies, 11 of the 16 dormitories were single-sex. This meant that in the tracking period, male and female students with the same high school graduation test scores in general lived in different dormitories. Under both policies, students lived in dormitories for at most two years before moving into private accommodation or university-owned apartments off campus. Hence, in 2006, the dormitory system contained both randomly assigned first year students and tracked second year students, who continued to live in the dormitory to which they were originally assigned. This group of first year students were not, therefore, exposed to the same mixing policy as those in 2007 and 2008 and so I omit 2006 from my main sample.¹¹

Quantifying the extent of the policy change: Figure 1.1 depicts the change in the composition of students' peer groups induced by the change in dormitory assignment policy. This figure shows the difference in mean peer group high school graduation test scores between the tracking and random assignment periods, for students at each percentile of the high school graduation test score distribution.¹² Students in the bottom decile lived with peers whose average high school graduation test scores were 0.5 standard deviations lower under tracking than random assignment. Students in the top decile lived with peers whose average high school graduation test scores were 0.4 standard deviations higher under tracking than random assignment. This provides a simple measure of the change in peer group academic proficiency (measured by high school graduation test scores) for different groups of students. The fact that observationally similar students had very different peer groups under the two policies provides the identifying variation I use to estimate the treatment effect of tracking.¹³

¹¹My results are robust to including this year in the random assignment period.

¹²I standardize graduation test scores to have a mean of zero and variance of one within each year in the control group.

¹³I create the figure in four steps. First, I calculate the mean high school graduation test score in each dormitory and assign this to each student as a measure of their peers' baseline academic proficiency. Second, I estimate a nonparametric regression of peers' baseline proficiency against own high school graduation test scores, separately for the tracking and random assignment period. I use a local linear regression with an Epanechnikov kernel and a plug-in bandwidth following Fan and Gijbels (1996). I allow the bandwidth to differ for the tracking and random assignment periods. Third, I calculate the vertical distance between the two fitted curves at each percentile of the distribution of high school graduation test scores. Finally, I use a nonparametric bootstrap to construct the 95% confidence intervals. The bootstrap resamples dormitory-year clusters 1000 times, stratifying by tracking/random assignment period, orders the resultant bootstrap estimates from smallest to largest and computes the confidence interval as the difference between the 25th and 975th estimates. The ordering is implemented separately for each point on the curve so the confidence

Figure 1.2 provides an alternative depiction of the effect of the policy on the distribution of student characteristics across dormitories. This figure shows the mean and standard deviation of high school graduation test scores in each dormitory-year unit. The tracking policy more than doubled the range of dormitory means from $(-0.2, 0.6)$ to $(-0.8, 12)$. The within-dormitory correlation coefficient rises from 0.12 under random assignment to 0.36 under tracking, which also emphasizes the substantial effect of the policy change on the distribution of students across dormitories. The within-dormitory correlation coefficient under tracking is still substantially less than one because of the *ad hoc* assignment of wait-listed applicants, race-specific assignment thresholds, and long lower tail of the graduation test score distribution (clearly visible in figure 1.3). This long tail creates a relatively high within-dormitory variance in “low-track” dormitories.

The figure also highlights some of the departures from pure tracking. First, the dormitory means under tracking are often relatively close together, because there are often “paired” male and female dormitories with similar scores. Second, the four observations in the left of the figure are clear outliers under both assignment policies. This is the university’s only self-catering dormitory, in which students purchase and prepare their own food rather than eating in communal dining halls. Students are permitted to request assignment to this dormitory and pay approximately 40% lower fees than in other dormitories. In practice it contains disproportionately many black students (who are likely to come from low-income households) and students with low high school graduation test scores. My results are entirely robust to excluding students in this dormitory, who make up only 1.4% of the sample.

Research design: My basic empirical strategy uses a linear difference-in-differences model to compare the GPA of dormitory students under the tracking policy and the random assignment policy, with non-dormitory students employed as a control group to remove any time trends in student GPAs. This strategy identifies the average treatment effect on the treated (ATT) of one policy relative to the other, assuming that the time trends in student GPAs for the treatment group (dormitory students) and control group (non-dormitory students) would have been identical in the absence of a policy change. I treat random assignment as the default policy and so define the parameter of interest as the

intervals should be interpreted as providing only pointwise coverage.

average treatment effect of tracking on the treated students:

$$\begin{aligned}\Delta^{ATT} &= \mathbb{E}[GPA(1)|D = 1, T = 1, X] - \mathbb{E}[GPA(0)|D = 1, T = 1, X] \\ &= \mathbb{E}[GPA(1)|D = 1, T = 1, X] - \mathbb{E}[GPA(0)|D = 1, T = 0, X] \\ &\quad - (\mathbb{E}[GPA(0)|D = 0, T = 1, X] - \mathbb{E}[GPA(0)|D = 0, T = 0, X]),\end{aligned}\tag{1.1}$$

where $GPA(1)$ and $GPA(0)$ respectively denote GPA for students who are and are not tracked, $D = 1$ for dormitory students, $T = 1$ for the period in which the tracking policy was in place, and X is a vector of students' demographic characteristics and high school graduation test scores. The second equality follows from the assumption of identical trends. I could alternatively define tracking as the default policy and estimate the treatment effect of random assignment - the results are numerically invariant to this choice. Note that difference-in-differences designs estimate “treatment on the treated” parameters, which are valid only for the treatment group (dormitory students) and do not necessarily speak to the effect that treatment would have on the control group (non-dormitory students).

I estimate Δ^{ATT} in the following regression model

$$GPA = \alpha + \beta D + \gamma T + \Delta TD + f(X) + \epsilon\tag{1.2}$$

where $f(\cdot)$ is a flexible function of the covariates. In practice, my results are robust to a wide range of different specifications. $\hat{\Delta}$ is a consistent estimator of Δ^{ATT} provided $f(\cdot)$ is correctly specified¹⁴ and the mean change in unobserved determinants of GPA from the random assignment to the tracking period is identical for dormitory and non-dormitory students:

$$\mathbb{E}[\epsilon|D = 1, T = 1, X] - \mathbb{E}[\epsilon|D = 1, T = 0, X] = \mathbb{E}[\epsilon|D = 0, T = 1, X] - \mathbb{E}[\epsilon|D = 0, T = 0, X].\tag{1.3}$$

This condition is not directly testable but I argue that it is plausible. The remainder of this

¹⁴I also estimate equation (1.2) adjusting for covariates using propensity score reweighting, following Abadie (2005), and using both reweighting and regression. These estimators may be more robust to misspecification of $f(\cdot)$ than the “regression-only” estimator (Robins and Rotnitzky, 1995). The estimates using regression, reweighting, and both regression and reweighting are all very similar.

section presents four arguments in support of this assumption.

First, table 1.1 shows the results of testing the assumption of equal time trends in students' baseline academic and demographic characteristics. The first panel shows that the time trends in mean high school graduation test scores are equal for dormitory and non-dormitory students. The same condition holds for the proportion of students scoring mostly As and mostly Ds in their high school graduation examinations. This rules out one *a priori* plausible violation of the assumption in equation (1.3): that the dormitory system might attract the more high achieving students in the tracking period than the random assignment period. The second panel shows that the time trends in the race, gender, and nationality composition of the student body were all approximately equal. The only characteristic that clearly violates assumption (1.3) is language, as the proportion of English-speaking students in the dormitories is higher in the tracking than the random assignment period.

Panel C shows that the probability of starting at the university in the random assignment period (2007 or 2008) after finishing high school early enough (2005 or earlier) to start during the tracking period is equal for dormitory and non-dormitory students. This suggests that students do not strategically delay their entrance to the university in order to manipulate the policy under which they are assigned. There is also no sign of such selection when restricting the test only to students near the top or bottom of the distribution of high school graduation test scores. These results show that the assumption of equal trends amongst treatment and control groups holds for students' observed characteristics and provides reassurance that the assumption may also hold for students' unobserved characteristics. I explore the consequences of violations of this assumption in appendix 1.A.4.

Second, neither assignment policy was well-publicized and little public information was made available about the policy change. The assignment policy in place at the time was not stated in the application and promotion materials distributed to potential students. Individual campus visits by prospective students are rare, so most students' only interaction with the university was through promotional materials, presentations by admissions staff at their high school, or informal interaction with current students. The assignment policy was available on the university's website, but this was one brief part of a lengthy document containing all information about student housing. The change in policy was not announced

in university or external media. Informal discussion with students suggested that relatively few were aware of the assignment policies when they started their studies and those who were aware tended to have older siblings or friends at the university.

Third, admissions staff report that students typically live in dormitories if and only if their family's permanent residence is outside Cape Town. The dormitory system is somewhat oversubscribed and so at most 4% of dormitory places are available for students from Cape Town and its suburbs. Other local students tend to live with their families, as the university is within one hour drive of most parts of the greater Cape Town metropolitan area. Students from outside Cape Town could, in principle, rent private accommodation rather than live in dormitories. However, private accommodation is typically considerably more expensive: a single room and three meals a day in a university dormitory in 2010 cost approximately 85% of the rent for a single room in a shared apartment equivalently close to the university.¹⁵ This suggests that students have relatively little discretion over whether they live in a dormitory or in private accommodation. Panel D of table 1.1 verifies this fact: the proportion of students from schools inside Cape Town who live in the dormitory system is approximately equal during the tracking and random assignment periods. The same condition holds for the proportion of students from schools outside Cape Town who do not live in the dormitory system. These equal time trends also hold for students at the top and bottom of the high school graduation test score distribution. This again suggests that students' decisions to enter the dormitory system were determined by where they lived, rather than strategic considerations.¹⁶

Fourth, there were no simultaneous policy changes at the university that are likely to affect the composition of the student body or their academic performance. The only change to the administration of the dormitory system was the closure of one old dormitory in 2006 and the opening of a new dormitory in 2007. My results are robust to omitting students in those dormitories. The admissions and financial aid policies at the university stayed constant

¹⁵The apartment rental rate is the average of the first ten apartments listed on the website www.gumtree.co.za on 30 October 2011 in the same neighborhood as the university.

¹⁶There are some students from Cape Town high schools who do live in dormitories and *vice versa*. This arises because dormitory admissions decisions are based on home location and this is imperfectly proxied by school location: some students at Cape Town schools live outside Cape Town and *vice versa*. Anecdotally, these students report either commuting long distances or living in hostels or dormitories during high school.

over this period. The tuition fee policy did change between 2005 and 2006: students were previously charged a flat rate for enrollment and this was replaced by a fee-per-credit hour system. I show in appendix 1.A.1 that the number of courses students took did not change over this period so there is little evidence that this fee policy change influenced students' behavior. This is unsurprising, as students were still required to enroll full time and the first year curriculum at this university requires a fairly standard number of courses. Inspection of minutes of Senate and other academic committee meetings shows no evidence of changes in grading standards or criteria at this time.

This discussion suggests that the “equal trends” assumption required for validity of the difference-in-differences design is plausible. In particular, the institutional setting allows limited scope for students to select strategically whether to live in dormitories in responses to the policy change: the change was not widely advertised and students' home location largely determined their admission to the dormitory system. Consistent with this claim, the time trends in observed characteristics for dormitory and non-dormitory students are approximately equal. Finally, there is little evidence of substantial changes in other policies at the university over this time period. The appendices also present detailed robustness checks that demonstrate that my results are robust to failures of the identifying assumption.

1.3 Treatment Effects of Tracking

I begin by estimating the average treatment effect of tracking on the treated students. The direction of this effect is not *a priori* clear: if tracking raises strong students' GPAs and lowers those of weak students, the sign of the average effect is ambiguous. I show that the data point to a large negative effect of tracking and then go on to explore how this effect varies throughout the distribution of student performance.

Average treatment effect of tracking: Table 1.2 reports estimates of the average treatment effect of tracking from the difference-in-differences model in equation (1.2). Column (1) reports a treatment effect of -0.13 standard deviations of GPA without conditioning on students' demographic characteristics or high school graduation test scores.¹⁷ The point

¹⁷I standardize GPA by subtracting the control group mean and dividing by the control group standard deviation in each year.

estimate is significant at the 10% level, with standard errors estimated using a cluster-robust variance matrix estimator that allows unrestricted correlation in the variances of students' unobserved characteristics at the dormitory-year level. Adding dormitory fixed effects to remove any time-invariant characteristics of the dormitories that might affect students' GPAs (such as access to study facilities, proximity to campus, shared rooms) reduces the treatment effect to a more precisely estimated -0.11 standard deviations. Column (3) adds a flexible set of conditioning variables: student gender, language, nationality, race, a cubic in high school graduation test scores and all possible two- and three-way interactions between these variables. Column (4) conditions on dormitory fixed effects and student characteristics, yielding a precisely estimated treatment effect of -0.12 standard deviations. Taken together, these results suggest that tracking reduced dormitory students' GPAs by approximately 0.12 standard deviations with a 95% confidence interval from 0.06 to 0.18 standard deviations, and that this effect is not explained by baseline characteristics of the students or dormitories.

I measure effect sizes in standard deviations of grade point average. This GPA measure is constructed from raw scores students obtain in each class, which range from 0 to 100. I take a credit-weighted average of each student's class-specific scores and then standardize the resultant student grades to have mean zero and standard deviation one in each year for the control group of non-dormitory students.¹⁸ The pre-standardized GPA measure has mean 58.9 with percentiles 1 and 99 equal to 7 and 83.6 respectively. No student ever obtains the maximum feasible grade of 100, so ceiling effects on this outcome measure are less of a concern in this setting than in US universities where many students obtain the maximum possible GPA. I use the raw numerical scores students obtain in each class, which are unlikely to have been "curved."

The effect size of 0.12 standard deviations is equal to approximately 150% of the male-female GPA gap and 30% of the black-white GPA gap in this sample of university students. Given the salience of race in the South African context, this points to a large and econom-

¹⁸Specifically, I first observe raw scores from 0 to 100 for every class that a student takes in her first year at the university. Second, I construct credit weightings for each class. Most of these weightings are recorded in the electronic dataset I obtained from the University of Cape Town; I construct the remaining weightings from physical copies of course catalogues from the relevant period. Third, I construct a credit-weighted average grade that ranges from 0 to 100 for each student in their first year at the university. Finally, I standardize this GPA measure to have mean zero and standard deviation one in each year for the control group of non-dormitory students.

ically meaningful treatment effect. Benchmarking the magnitude of this effect relative to the existing literature is complicated by the fact that there are few prior studies of different peer group assignment rules. The two most similar studies in the existing literature yield comparably sized treatment effects, assuming that standard deviation-based scales are commensurable across such different settings. First, Duflo, Dupas, and Kremer (2011) find that instructional tracking in Kenyan primary schools increases students’ test scores by 0.14 to 0.18 standard deviations relative to random assignment. Second, Carrell, Sacerdote, and West (2012) find that their “optimal” rules for assigning students to noninstructional peer groups at the US Air Force Academy reduce students’ GPAs by 0.08 to 0.1 standard deviations relative to random assignment. These effects are small relative to many of the resource- and instruction-based interventions in developing countries reviewed in Glewwe and Kremer (2006). However, the low cost involved in redistributing students across groups may make this type of intervention more attractive to policymakers.

How does tracking affect different subgroups? This subsection disaggregates the negative average treatment effect of tracking using a variety of heterogeneous treatment effect estimators. These estimates consistently show large negative treatment effects near the bottom of the GPA distribution and approximately zero treatment effects on students near the top of the GPA distribution.

Table 1.3 reports estimates of the average treatment effect on the treated for each quartile of the distribution of high school graduation test scores. The effects are in fact negative in all four quartiles but are considerably larger below the median (-0.22 to -0.26 standard deviations) than in the top quartile (0.09 standard deviations). These results clearly point to negative effects on weaker students but suggest that even those students with high baseline performance might have lower GPAs under tracking.

It is, however, possible that quartiles are too coarse a division and that there are some subgroups of students in the top quartile who do benefit from tracking. I explore this possibility by estimating a local linear regression of GPA on high school graduation test scores, separately for dormitory students in the tracking period, dormitory students in the random assignment period, non-dormitory students in the tracking period, and non-dormitory students in the random assignment period. I then use the local linear estimates $\hat{GPA}_{DT}(HS)$

for group D in period T to compute the local difference-in-differences estimator

$$\hat{\Delta}^{ATT}(HS) = G\hat{P}A_{11}(HS) - G\hat{P}A_{10}(HS) - G\hat{P}A_{01}(HS) + G\hat{P}A_{00}(HS)$$

at each percentile of high school graduation test scores (HS). This provides a flexible test of whether tracking has a positive effect for any subset of students. Panel A of figure 1.4 plots $\hat{\Delta}^{ATT}(HS)$ and shows that the point estimates are negative for approximately 80% of the distribution of high school graduation test scores.¹⁹ The treatment effect is equal to approximately one quarter of a standard deviation in much of the bottom quartile, while the positive effect in the left tail never exceeds one tenth of a standard deviation. The bootstrap confidence intervals are relatively wide but the estimates are negative and significant from approximately the 15th to the 65th percentiles.²⁰ This nonparametric regression does not control for student or dormitory characteristics but results are robust to splitting the sample and estimating $\hat{\tau}^{ATT}(HS)$ separately by octile with the same conditioning variables used in the previous section.²¹

Table 1.4 reports average treatment effects on the treated for several demographic subgroups. The treatment effects on male and female students are equal, but black students experience a substantially larger negative effect than white students (-0.23 standard deviations compared to -0.14 standard deviations). This difference is not significant but it suggests that black students, who on average have lower high school graduation test scores and lower socio-economic status, are disproportionately affected by the tracking policy. There is a positive but entirely insignificant treatment effect on students of other race groups; the rea-

¹⁹I use an Epanechnikov kernel and a plug-in bandwidth chosen separately for each of the four nonparametric regressions (Fan and Gijbels, 1996). The results are qualitatively robust to substantial changes in the bandwidth parameter.

²⁰This bootstrap algorithm resamples dormitory-year clusters, computes the local linear estimate for the resampled data, orders the resultant point estimates from 1000 replications, and computes the confidence intervals as the difference between the 25th and 975th estimates. The ordering is implemented separately for each percentile of the distribution of high school graduation test scores so the confidence intervals should be interpreted as providing only pointwise coverage. These confidence intervals depend on non-pivotal statistics and so do not offer asymptotic refinement but do avoid the need to rely on critical values from the standard normal distribution that may be a poor approximation to the finite sample distribution of the test statistic (Horowitz, 2001).

²¹Including controls in the local linear model is in principle possible. However, existing estimators for this “partially linear” model either assume that the distribution of high school points is strictly continuous (Yatchew, 1997) or require the choice of multiple bandwidth parameters and are highly sensitive to these choices (Robinson, 1988).

son for this effect is unclear. Given that black students have on average lower high school graduation test scores than white students, these results are consistent with the idea that lower-scoring students are affected more by tracking than higher-scoring students.

Quantile treatment effects of tracking: The nonlinear difference-in-differences model proposed by Athey and Imbens (2006) provides an alternative way to explore heterogeneous treatment effects, which I discuss in detail in appendix 1.B. In brief, the model constructs the full counterfactual distribution (CDF) of GPAs for dormitory students under tracking *if tracking had not been implemented*, whereas the standard linear difference-in-differences model constructs only the counterfactual mean of this distribution. The observed and counterfactual distributions of graduation test scores are shown in figure 1.5. The horizontal difference between these two distributions at each quantile is defined as the *quantile treatment effect on the treated* and these are plotted for each quantile in the first panel of figure 1.6. These are large and negative in the lower tail of the distribution, with effects of more than -0.6 standard deviations at the 5th percentile and more than -0.3 standard deviations at the 10th. The effects are smaller at higher percentiles but remain negative for approximately 90% of the distribution. The estimates are relatively imprecisely estimated and statistically differ from zero only in the lowest quartile of the distribution²² but the results clearly reaffirm that tracking has a substantial negative effect on the lower tail of the distribution and little or no effect on the upper tail.²³

The additional flexibility of the nonlinear difference-in-differences model comes at the cost of a stronger identifying assumption: that the distribution of baseline characteristics is constant through time for each group (dormitory and non-dormitory students). I show in appendix 1.B that this restriction can be relaxed to allow for changes in observed student characteristics through time. This relaxation operates by reweighting students to equalize

²²I again construct the confidence intervals using the percentile cluster bootstrap algorithm discussed in footnote 20. The validity of the bootstrap for the nonlinear difference-in-differences estimator has not been formally established. However, Athey and Imbens (2006) present simulation results showing that tests based on bootstrap confidence intervals have coverage probabilities closer to their nominal levels than confidence intervals based on the analytical variance estimator that they derive.

²³Note that these are treatment effects on quantiles of the outcome distribution, not treatment effects on individual students. The two concepts are equivalent only if tracking is a rank-preserving treatment, so that the students in the bottom of the observed distribution would also be at the bottom of the counterfactual distribution (Heckman, Smith, and Clements, 1997).

the distribution of observed baseline characteristics through time, following DiNardo, Fortin, and Lemieux (1996) and Hirano, Imbens, and Ridder (2003). The second panel of figure 1.6 shows the quantile treatment effects estimated by this reweighted nonlinear difference-in-differences estimator. The results are very similar to those from the unadjusted estimator, reinforcing the ongoing theme that the effects of tracking are not confounded by differences in students' observed characteristics between the tracking and random assignment periods.

Figures 1.4 and 1.6 convey complementary but subtly different information about the nature of the treatment effects. The former figure presents *treatment effects for students at different points in the distribution of high school graduation test scores*. This shows that students with low high school graduation test scores were significantly harmed by the policy and that students with high graduation test scores may have been helped slightly. The latter figure presents *treatment effects at different points in the distribution of university GPA*. This shows that the lower tail of the GPA distribution dropped by a very large margin under tracking and the upper tail was unaffected. These latter statements are about the distribution of outcomes and make no claim about which students are at which point of this distribution. The former statements apply to specific subgroups of students defined by their high school graduation test scores. The two analyses present a consistent picture – tracking hurts weaker students and does little to help strong students – but do so using different methods.

Having constructed the full counterfactual distribution of GPAs, I can also compare a range of summary statistics for the observed and counterfactual distributions. In particular, the first row of table 1.5 shows that the average treatment effect on the treated from the nonlinear model is -0.13 standard deviations, or -0.1 standard deviations after conditioning on student characteristics and dormitory fixed effects. This is very similar to the treatment effect estimated by the linear difference-in-differences model, which provides some reassurance that my results are not driven by differences in the two models' assumptions.

Inequality treatment effects of tracking: The next four rows of table 1.5 show that standard measures of inequality in academic performance are sharply higher for the distribution of tracked outcomes than the counterfactual distribution. The interdecile range rose by approximately 15%, the interquartile range by approximately 10%, the Gini coefficient

by approximately 20% and the coefficient of variation by a considerably smaller margin.²⁴ These differences are all significant and suggest that any social welfare function that values both average GPA and equality of GPA would find tracking a particularly unattractive policy in this setting.²⁵ This finding is also relevant for the academic tracking literature more broadly. This literature strongly emphasizes inequality considerations (see Betts, 2011) but few papers quantify the effect of tracking on standard inequality measures. I provide a simple method of quantifying this effect, which improves on the existing approaches of estimating average subgroup treatment effects and then informally discussing whether these suggest increases in inequality.

Mobility treatment effects of tracking: The previous subsections have established that tracking has a large negative effect on low-scoring students and at most a small positive effect on high-scoring students. This raises the concern that tracking might also entrench the baseline ranking of academic performance, by impairing the potential for students who begin university near the bottom of the distribution of academic performance to catch up to or overtake their peers. This would concern any policymaker or social planner who values academic mobility and the capacity of university-level policy to reduce or reinforce differences in academic performance relative to high school. Given that the South African education system is characterized by high levels of socio-economic inequality (as are many others in developed and developing countries), this concern is highly salient.

This subsection therefore examines the relative effects of tracking and random assignment to dormitories on changes in students' rank mobility between high school and university. Specifically, I test whether the two policies have different effects on the probability that students will change their rank in the distribution of academic performance from high school

²⁴The Gini coefficient is only defined for random variables with strictly positive support. I therefore add 4 to the standardized GPA measure to ensure that all values are positive. The treatment effect on the Gini coefficient, expressed in percentage terms, is robust to alternative location shifts provided these ensure a strictly positive support. This also ensures that the mean is non-zero, which is necessary for estimating the coefficient of variation.

²⁵Note that I construct the counterfactual distributions quantile-by-quantile so the interdecile and interquartile ranges are direct by-products of this process but the mean, Gini coefficient, and coefficient of variation require linear approximation between the quantiles. This linear approximation introduces non-classical measurement error into the estimation so the estimates should be interpreted with a degree of caution. I attempt to minimize the extent of the error by estimating the counterfactual distribution at 199 quantiles.

to university. This complements the analyses in the previous section, which showed that tracking had a significant negative effect on GPAs in the lower tail of the distribution. This *level effect* demonstrates that random assignment helps low-achieving students to “catch up” to their peers. The *rank analysis* I present in this section demonstrates that random assignment is not sufficient to facilitate “overtaking” or rank changes between high school and university.

I begin by constructing transition matrices for the tracking and random assignment periods. Each p_{ij} element of these four-by-four matrices indicates the probability that an individual in quartile i of the distribution of high school graduation test scores will move to quartile j of the distribution of university GPA. A diagonal matrix corresponds to zero mobility (every student remains in the same quartile), while $p_{ij} = 0.25 \forall i, j$ corresponds to complete mobility (students’ final quartiles are independent of their initial quartile). I pool dormitory and non-dormitory students’ graduation test scores and GPAs to compute the quartile boundaries but present the transition probabilities for only dormitory students.

The two transition matrices are reported in table 1.6. The first row of panel A shows that dormitory students in the tracking period who are in the bottom quartile of the high school graduation test score distribution remain in the bottom quartile in university with probability 0.37 and move to the top quartile with probability 0.28. A visual inspection of panel A (tracking) and panel B (random assignment) suggests that the policy change had little effect on the transition probabilities.

Table 1.7 presents several summary measures of mobility in each period. I define a *mobility treatment effect* as the change in a summary mobility measure from the random assignment to the tracking period. The first row shows that the average probability that a student will move from one quartile to another is slightly higher than 0.8 in each period (Bartholomew, 1982). The average number of quartile changes is approximately 0.9 in each period (Bartholomew, 1982) and the correlation between initial and final quartiles (measured by the second largest eigenvalue of each transition matrix) is approximately 0.55 (Sommers and Conlisk, 1979). None of the mobility treatment effects are significant and the magnitudes are very small, suggesting that the policy change raised the level of low-achieving students’ GPAs but had little effect on their probability of overtaking students with higher graduation

test scores in high school.

1.4 What Can Be Learned from Cross-Dormitory Variation?

The preceding section presents indirect evidence of the importance of peer effects in determining students' GPAs. My argument can be characterized as follows: students' GPAs changed sharply when the residential tracking policy was introduced, this change cannot be explained by differences in student or dormitory characteristics, no substantial simultaneous policy changes occurred, so the change must be due to the different peer groups created by the tracking policy. This section complements the argument by presenting a direct test for peer effects using cross-dormitory variation in peer characteristics and then exploring the structure of the education production function.

Begin by considering a simple model of students' GPA production adapted from Manski (1993) and Sacerdote (2001):

$$GPA_{ig} = \alpha_0 + \alpha_1 HS_{ig} + \alpha_2 \overline{HS}_g + \alpha_3 \overline{GPA}_g + \epsilon_{ig}, \quad (1.4)$$

where GPA_{ig} and HS_{ig} are the university GPA and high school graduation test score respectively of student i in dormitory g . Define \overline{GPA}_g and \overline{HS}_g as the average GPA and high school graduation test score of all students in dormitory g . In the language of Manski (1993), α_2 is an "exogenous peer effect," in which students' GPAs are influenced by the baseline characteristics of their peers and α_3 is an "endogenous peer effect," representing a feedback loop between each students' GPA and that of her peers. It is clearly impossible to estimate equation (1.4) consistently, as the same variable appears on both sides of the equation. However, it is possible to evaluate equation (1.4) at the dormitory average, solve for

$$\overline{GPA}_g = \frac{\alpha_0}{1 - \alpha_3} + \frac{\alpha_1 + \alpha_2}{1 - \alpha_3} \overline{HS}_g + \frac{1}{1 - \alpha_3} \bar{\epsilon}_g$$

and substitute this back into equation (1.4) to obtain the reduced form

$$\begin{aligned} GPA_{ig} &= \frac{\alpha_0}{1 - \alpha_3} + \alpha_1 HS_{ig} + \frac{\alpha_2 + \alpha_1 \alpha_3}{1 - \alpha_3} \overline{HS}_g + \epsilon_{ig} + \frac{\alpha_3}{1 - \alpha_3} \bar{\epsilon}_g \\ &\equiv \pi_0 + \pi_1 HS_{ig} + \pi_2 \overline{HS}_g + \eta_{ig}. \end{aligned} \tag{1.5}$$

If students are randomly assigned to groups, \overline{HS}_g is uncorrelated with η_i , which is a student-specific deviation from the average dormitory unobserved characteristics. Then this model can be consistently estimated and $\pi_2 \neq 0$ if and only if some peer effect exists: $\alpha_2 \neq 0$ or $\alpha_3 \neq 0$.²⁶ Sacerdote (2001) therefore proposes $H_0 : \hat{\pi}_2 = 0$ as a reduced-form test for the existence of either peer effect.

Table 1.8 reports the results of this test for the sample of all dormitory students who are randomly assigned to dormitories. The value of $\hat{\pi}_2 \approx 0.24$ implies that a one standard deviation increase in peers' high school graduation test scores is associated with a one quarter standard deviation increase in each student's GPA in their first year of university. This result is robust to conditioning on students' demographic characteristics (column 2), to including dormitory fixed effects (column 3), and to excluding the one outlying dormitory discussed in section 1.2 (columns 4–6).²⁷

While it is impossible to recover the values of the structural parameters α_2 and α_3 from the reduced form coefficients π_1 and π_2 , it is worth noting that if $\alpha_3 = 0$, then $\hat{\alpha}_2 = \hat{\pi}_2 \approx 0.24$ (standard error 0.09) and if $\alpha_2 = 0$, then $\hat{\alpha}_3 = \frac{\hat{\pi}_2}{\hat{\pi}_1 + \hat{\pi}_2} \approx 0.4$ (standard error 0.06). The latter estimate of the endogenous peer effect or social multiplier implies that approximately two fifths of the gain from a dormitory-level increase in GPA spills over onto each student. These effect sizes are comparable to the larger results from previous studies of residential peer effects in higher education. This may reflect the fact that my measure of peer academic proficiency, standardized scores on content-based high school graduation tests, differs from the scholastic aptitude test (SAT) scores used in most prior studies. The latter measure assigns a relatively higher weight to peers' academic ability and less to their study behavior and content knowledge. Stinebrickner and Stinebrickner (2006) present evidence suggesting

²⁶Unless one peer effect is negative, a possibility seldom considered in the literature.

²⁷I also estimate equation (1.5) using dormitory-by-year medians instead of means. The point estimates are very similar though slightly less precisely estimated.

peer effects are larger when using measures of peer characteristics that attach more weight to study behavior and knowledge.

Note that the variation in \overline{HS}_g used to estimate $\hat{\pi}_2$ arises because randomization balances dormitory means only in expectation, not in any particular finite sample. This means that the regressor of interest takes on only 30 values, which are relatively close together. This narrow support of the regressor and the cluster-robust variance estimator I use yield relatively large standard errors. The point estimate suggests that a one standard deviation increase in peers' mean graduation test scores in high school increase own GPA by one quarter of a standard deviation. However, the range of \overline{HS}_g is only 0.85 (after excluding the outlying dormitory), so moving from the "worst" to the "best" observed peer group would raise GPA by one fifth of a standard deviation.

The linear-in-means model is a standard point of departure in the empirical peer effects literature but it embodies a number of important limitations. In particular, the model requires that own and mean peer group high school graduation test scores are additively separable in the production function. Moving a high-scoring student from one dormitory to another will increase average student performance in the second group by exactly the same amount that it decreases average student performance in the first group. Hence, any reallocation of students between peer groups will leave the average GPA unchanged. This implies a zero average treatment effect of tracking, as well as any other reallocation policy. Combining this observation with the empirical results in section 1.3 suggests that the linear-in-means model is seriously misspecified.

I therefore estimate a more general model of the form

$$GPA_{ig} = f(HS_{ig}, \overline{HS}_g) + X_{ig}\Gamma + \eta_{ig} \quad (1.6)$$

where $f(\cdot, \cdot)$ is a polynomial function and X_{ig} is a vector of student characteristics and dormitory fixed effects. I estimate f using a series estimator with the order of the polynomial chosen by cross-validation.²⁸ A polynomial order of 2 minimizes of the mean squared error

²⁸I use a leave-out-one-cluster cross validation scheme that allows for possible dependence of the error structures within dormitory-year clusters. This follows the spirit of the literature on cross-validation for dependent data in a time series context (Burman, Chow, and Nolan, 1994). The quadratic model is also the

of the regression, so I estimate the model

$$GPA_{ig} = \beta_0 + \beta_1 HS_{ig} + \beta_{11} HS_{ig}^2 + \beta_2 \overline{HS}_g + \beta_{22} \overline{HS}_g^2 + \beta_{12} HS_{ig} \overline{HS}_g + X_{ig} \Gamma + \eta_{ig} \quad (1.7)$$

with and without the vector of student demographic characteristics X_{ig} .²⁹

Table 1.9 reports estimates of the parameters of the augmented model (1.7). The key results are that $\hat{\beta}_{12}$ is consistently negative across all specifications, while the sign and magnitude of $\hat{\beta}_{22}$ are somewhat sensitive to specification of the control vector. A likelihood ratio test strongly rejects the linear model relative to the quadratic model. The negative value of $\hat{\beta}_{12}$ indicates that students with low high school graduation test scores benefit more from an increase in mean peer high school graduation test scores than do students with higher graduation test scores. This implies that tracking, by blocking interaction between low and high scoring students, will hurt the former group more than it helps the latter and so reduce average GPA. The negative average treatment effect of tracking estimated in section 1.3 is entirely consistent with this result. Both the theoretical and empirical literature on peer effects (and neighborhood segregation) have emphasized the importance of this complementarity parameter and my results are consistent with the claim that when own and peer characteristics are partially substitutable, tracking reduces mean outcomes.

The inconclusive sign of $\hat{\beta}_{22}$ is also of interest. If this parameter is negative, then GPA is a concave function of peers' high school graduation test scores. Average output will therefore be lower with one high and low scoring peer group than with two groups with equal mean high school scores.³⁰ Unlike $\hat{\beta}_{12}$, this parameter does not provide any information about which students benefit most from strong peers. However, it does provide important information about whether mixed or tracked peer groups maximize average output. The role of convexity or concavity in peer effects models has received relatively little attention in the empirical literature but is emphasized in theoretical and methodological work by Benabou (1996) and Graham, Imbens, and Ridder (2011). If both $\hat{\beta}_{12}$ and $\hat{\beta}_{22}$ were negative, the negative average

specification that minimizes the Bayesian Information Criterion for this model. Note that this estimation strategy departs from a fully nonparametric approach by assuming that X_{ig} is linear and separable from f .

²⁹I also estimate this model using dormitory-by-year medians instead of means. The point estimates are very similar though slightly less precisely estimated.

³⁰This follows from Jensen's inequality: for a concave function $g(\cdot)$, $\frac{1}{2}g(\mu_{low}) + \frac{1}{2}g(\mu_{high}) \leq g\left(\frac{\mu_{low} + \mu_{high}}{2}\right)$.

effect of tracking would have been predictable using cross-dormitory variation. Given the ambiguous sign of $\hat{\beta}_{22}$, out-of-sample prediction might lead to incorrect conclusions.

Even if estimates of the two key parameters were consistently negative, predicting the negative effect of tracking would require an ambitious out-of-sample extrapolation. Figure 1.7 panel A shows the density of dormitory-level mean high school graduation test scores under tracking and random assignment. Tracking generates a much more dispersed distribution of means ranging from -0.8 to 1.2 standard deviations, while those under random assignment range from -0.2 to 0.6. A similar problem applies to other dormitory-level statistics such as the variance and proportion of students at different percentiles of the high school graduation test score distribution, shown in panels B – D. This figure highlights the limitation of existing methods for predicting the effect of changes in group assignment policies or inferring optimal assignment policies (Bhattacharya, 2009; Graham, Imbens, and Ridder, 2011). These methods cannot predict out of sample without strong assumptions and so cannot speak to the effect of changing the policies that create peer groups that are not observed under the status quo assignment policy. This may account for the puzzling result in Carrell, Sacerdote, and West (2012), where a change in group assignment policy reduced average outcomes instead of raising them, as estimates based on randomly assigned groups suggested. The “optimal” groups created by the new policy were not observed under the old assignment policy and so their estimated optimality relied on out-of-sample projection. A similar argument can be applied to aspects of the results in Duflo, Dupas, and Kremer (2011). They estimate variants of the linear-in-means model in equation (1.5) using students who are randomly assigned to first grade classrooms in Kenyan schools. They interpret this result as evidence that tracking students into classrooms should hurt low-scoring students and help high-scoring students, unless instructors adapt their behavior to respond to the changed classroom composition. However, it is also possible that their estimates should not be extrapolated out-of-sample to tracked classrooms.

1.5 Mechanisms

Having established the effects of tracking in section 1.3 and the effect of marginal changes in peer group composition in section 1.4, I now consider the mechanisms that might be driving

these results. In particular, I suggest that residential peer effects may operate through two conceptually distinct channels: a direct channel, in which peers exert influence through spatial proximity,³¹ and an indirect channel, in which spatial proximity influences the formation of social and/or academic networks, which in turn influence student outcomes. The models discussed in section 1.4 estimated a reduced form combination of these effects. If indirect effects are important, then changes in dormitory assignment policies might change the relationship between spatial proximity and social networks. Estimates that combine direct and indirect effects would then be sensitive to assignment policies and so could not reliably predict the effect of changes in these policies. For example, if students prefer to interact with peers of similar academic proficiency, then intra-dormitory social networks will be stronger under tracking than random assignment. The size of the indirect effects will therefore be magnified under tracking and peer effects estimated under random assignment will be understated. This further emphasizes the importance of explicit cross-policy comparisons, even if the out-of-sample extrapolation and model specification problems were not present.

I present evidence in this section that direct effects alone cannot explain the estimated peer effects. Specifically, I estimate an augmented version of the reduced-form linear-in-means model that allows peer effects to differ within and across race groups:

$$GPA_{i,g,r} = \psi_0 + \psi_1 HS_{i,g,r} + \psi_2 \overline{HS}_{g,r} + \psi_3 \overline{HS}_{g,-r} + \nu_{i,g,r}. \quad (1.8)$$

where $\overline{HS}_{g,r}$ is the average high school graduation test score for students in dormitory g of the same race as student i and $\overline{HS}_{g,-r}$ is the average for students of other races. Table 1.10 reports that for randomly assigned dormitory students, $\hat{\psi}_2 \approx 0.16$ (standard error 0.09) and $\hat{\psi}_3 \approx -0.05$ (standard error 0.07). These results suggest that peer effects occur almost entirely within race groups. Given the salience of race in contemporary South Africa, social and study networks may be strongly correlated with race groups. If this is true, the near-zero cross-race peer effects are evidence that spatial proximity (measured by assignment to the same dormitory) alone does not predict academic outcomes. Instead, the strong within-race peer effects suggest that spatial proximity matters by influencing students' networks, and

³¹For example, noisy neighbors in the dormitory might impair students' ability to study, even if no direct interaction occurs between the students.

hence their academic outcomes.

I estimate a similar variant of the linear-in-means model that allows the effect of change in peers' mean high school graduation test scores to differ for students in the same and different colleges/faculties/schools, such as engineering, science, and social science. Table 1.11 shows that peer effects are not stronger within than across faculties. The point estimates for within-faculty effects are in fact consistently smaller than those for cross-faculty effects, though I cannot reject equality of the coefficients in any specification.³² This suggests that peer effects do not primarily operate through direct academic collaboration (study groups, cooperation on essays, copying problem sets, etc.). Peer effects may instead be operating through time allocation toward studying and leisure activities, a result consistent with Stinebrickner and Stinebrickner (2008). However, without detailed time use data I cannot explicitly test this hypothesis.

An unusual feature of the assessment system at the University of Cape Town provides additional insight into the time at which peer effects matter. In particular, I show that the treatment effect operates early on in the semester, rather than being concentrated on students' performance in final examinations. Course-level assessment at many South African universities takes place in two stages. In the first stage, students are graded on their performance in tests, problem sets and class discussion, and sometimes their attendance record. Students who perform particularly poorly in this assessment stage may be refused permission to proceed to the second stage of assessment, typically a final examination. I refer to these as "excluded courses" and the transcript data that I use simply shows an exclusion symbol for these courses, rather than a numerical grade.

The results in section 1.3 assigned zero grades to excluded courses. I can instead drop these courses from the sample, calculate students' GPAs using only the non-excluded courses and estimate the treatment effects on this alternative outcome measure.³³ The third and

³²I obtain similar results when I estimate the effect on students' results in a particular class, such as Economics 1. There is no evidence that the characteristics of peers taking Economics 1 matter more than the characteristics of peers not taking the course. The same result holds for most large introductory classes.

³³Note that this is not a standard censored data problem. If a student is academically excluded from a course I still observe them in the data and they still obtain grades for all their other courses. Less than 0.05% of the sample are academically excluded from all the courses that they take in first year. This group of students enter the main results as having a zero GPA (before standardization) and are omitted from the regressions reported in columns (3) and (4) of table 1.12.

fourth columns of table 1.12 show that this reduces the average treatment effects on the treated to -0.07 standard deviations (standard error 0.03). This is approximately half as large as the treatment effect on the original GPA measure but is still significant. The fifth and sixth columns estimate the treatment effect on the proportion of courses for which students are excluded. This rises from 5 percentage points in the absence of tracking to 8 percentage points under tracking. A large part of the average treatment effect is therefore operating at the extensive margin, through the mechanism of course exclusions.

These additional outcome measures provide some important evidence about the mechanisms through which tracking is lowering GPA. The fact that course exclusions are based on performance early in the semester suggests that tracking is reducing students' academic aptitude or application from an early stage in the courses, with a particularly deleterious effect on the bottom tail of the distribution. This is more consistent with a model in which peer effects operate from early in the semester than a model in which peers only effect time allocation during intensive study periods just before final examinations.

1.6 Conclusion

This paper describes in detail the difference in student GPAs between a period in which students were tracked into dormitories and a period in which they were randomly assigned to dormitories at the University of Cape Town in South Africa. I show that under tracking the mean GPA was lower and the level of GPA inequality higher. This result arises because low-scoring students experience a large drop in academic performance when living with only low-scoring peers, while high-scoring students experience no effect on their academic performance when living with only high-scoring peers. These substantial peer effects occur largely from interaction with spatially proximate own-race peers and the relevant form of interaction does not appear to be direct academic collaboration. I present an extensive argument supporting a causal interpretation for these results.

These results demonstrate that different policies for assigning students to peer groups can have large effects on their academic performance. In particular, academic tracking may generate a substantially worse distribution of academic performance than random assignment. However, caution should be exercised in using my results to judge holistically the

relative merits of the two policies. Different assignment policies may entail transfers from some groups of students to others and, as academic outputs are typically non-tradeable, it may not be possible to Pareto rank different policies. Many non-measured student outcomes may also be affected by different group assignment policies, including time use (Stinebrickner and Stinebrickner, 2008) and attitudes toward diversity (Boisjoly, Duncan, Kremer, Levy, and Eccles, 2006). For example, high-scoring students' performance may be unaffected by tracking because the rise in their peers' academic proficiency induces them to substitute time away from studying toward leisure. In future work I plan to study the long-term effects of tracking versus random assignment on graduation rates, time-to-degree, and labor market outcomes. These results will permit a more comprehensive evaluation of the relative merits of the two group assignment policies.

Despite these concerns, my findings provide important evidence regarding the importance of peer group assignment policies. I provide what appears to be the first clean evidence on the effects of noninstructional tracking. This complements the small literature that cleanly identifies the effect of instructional tracking. For example, Duflo, Dupas, and Kremer (2011) find that although the total effect of instructional tracking is positive, there is suggestive evidence that this combines a negative direct peer effect of tracking with a positive effect due to changes in instructor behavior. My findings also suggest that policymakers can change the distribution of students' academic performance by rearranging the groups in which these students interact while leaving the marginal distribution of inputs into the education production function unchanged. This is attractive in any setting but particularly in resource-constrained developing countries. While the external validity of any result is always debatable, my findings may be particularly relevant to universities serving a diverse student body that includes both high performing and academically underprepared students. This is particularly relevant to selective universities with active affirmative action programs, such as those studied in Bertrand, Hanna, and Mullainathan (2010).

The examination of peer effects under random assignment in sections 1.4 and 1.5 also points to fruitful avenues for future research. These results show that peer effects estimated under random assignment had limited ability to predict the effects of a change in assignment policy and that peer effects do not operate only directly through spatial proximity but

instead appear to be mediated by changes in students' social networks. This highlights the risk of relying too heavily on reduced form estimates that do not explicitly acknowledge the behavioral content of peer effects. Research that combines peer effects estimated under different peer group assignment policies with detailed data on social interactions and explicit models of network formation may provide additional insights.

Table 1.1: Baseline demographics and high school graduation test scores for dormitory and non-dormitory students in the tracking and random assignment periods

| | (1) Dormitory students Tracked | (2) Randomized | (3) Other students Tracked | (4) Randomized | (5) Second difference |
|--|--------------------------------------|-------------------|----------------------------------|-------------------|-----------------------------|
| <u>Panel A: High school graduation test scores</u> | | | | | |
| Mean score (standardized) | .17 | .20 | .00 | .00 | -.03 (.04) |
| % “A” students | .20 | .21 | .15 | .18 | .02 (.01) |
| % “D” students | .10 | .10 | .12 | .13 | .01 (.01) |
| <u>Panel B: Demographic characteristics</u> | | | | | |
| % female | .50 | .52 | .52 | .51 | -.03 (.02) |
| % black | .50 | .52 | .12 | .12 | -.02 (.01) |
| % white | .35 | .33 | .52 | .49 | 0 (.02) |
| % other races | .13 | .13 | .34 | .37 | .03* (.01) |
| % English-speaking | .59 | .56 | .85 | .86 | .05*** (.01) |
| % international | .23 | .18 | .11 | .06 | 0 (.01) |
| <u>Panel C: Year of high school graduation</u> | | | | | |
| % who graduated in an early enough year to be eligible to enroll under tracking | | .027 | | .033 | .006 (.004) |
| % “A” students eligible for tracking | | .001 | | .003 | .002 (.003) |
| % “D” students eligible for tracking | | .055 | | .048 | -.007 (.016) |
| <u>Panel D: Geographic location of high school</u> | | | | | |
| In Cape Town | .116 | .106 | .770 | .749 | -.012 (.013) |
| % “A” students in Cape Town | .112 | .078 | .811 | .785 | .008 (.028) |
| % “D” students in Cape Town | .246 | .228 | .775 | .762 | .004 (.049) |

Notes: This table shows that the change in students’ baseline characteristics between the random assignment and tracking periods is approximately equal for dormitory and non-dormitory students. This is evidence that students are not strategically choosing whether to live in dormitories and provides support for the assumptions of the research design. Standard errors in parentheses are estimated using White’s heteroscedasticity-robust covariance matrix. ***, ** and * denote significance at the 1%, 5% and 10% levels respectively.

Table 1.2: Estimates of the average treatment effect of tracking using a linear difference-in-differences regression model

| | (1) | (2) | (3) | (4) |
|---|------------------|-------------------|--------------------|--------------------|
| Sample | Full sample | Restricted sample | | |
| Dependent variable | GPA | | | |
| Average treatment effect of tracking on the treated | -.129* (.073) | -.107* (.061) | -.111*** (.034) | -.123*** (.031) |
| Dormitory fixed effects | | × | | × |
| Individual controls | | | × | × |
| # dorm-year clusters | 60 | 60 | 60 | 60 |
| # dormitory students | 7480 | 7480 | 6600 | 6600 |
| # other students | 7188 | 7188 | 6685 | 6685 |

Notes: Standard errors in parentheses are estimated using a cluster-robust variance estimator that allows unrestricted error correlation at the dormitory-year level, with non-dormitory students treated as singleton clusters. ***, ** and * denote significance at 1%, 5% and 10% levels respectively. Restricted sample in columns (3) and (4) excludes observations with missing individual controls. Columns (3) and (4) control for student gender, race, language, nationality, a cubic in high school graduation test scores and all possible three-way interactions.

Table 1.3: Estimates of the average treatment effect of tracking for students in each quartile of the high school graduation test score distribution using a linear difference-in-differences regression model

| | (1) | (2) | (3) | (4) | (5) |
|---|--------------------|---|--------------------|--------------------|-----------------|
| Sample | All students | Quartiles of high school graduation test scores | | | |
| Dependent variable | | Fourth | Third | Second | First |
| | GPA | | | | |
| Average treatment effect of tracking on the treated | -.123*** (.031) | -.224*** (.075) | -.261*** (.058) | -.157*** (.048) | -.091* (.05) |
| Dormitory fixed effects | × | × | × | × | × |
| Individual controls | × | × | × | × | × |
| # dorm-year clusters | 60 | 59 | 59 | 58 | 57 |
| # dormitory students | 6600 | 1298 | 1301 | 1475 | 1811 |
| # other students | 6685 | 1715 | 1663 | 1663 | 1460 |
| F-test for equality of quartile effects | | | | | |
| without controls | | | | 3.706** | |
| with controls and fixed effects | | | | 1.85 | |

Notes: Standard errors in parentheses are estimated using a cluster-robust variance estimator that allows unrestricted error correlation at the dormitory-year level, with non-dormitory students treated as singleton clusters. ***, ** and * denote significance at 1%, 5% and 10% levels respectively. Observations with missing individual controls are excluded from the sample. All columns control for student gender, race, language, nationality, a cubic in high school graduation test scores and all possible three-way interactions.

Table 1.4: Estimates of the average treatment effect of tracking for race and gender subgroups using a linear difference-in-differences regression model

| Sample | (1) | (2) | (3) | (4) | (5) | (6) |
|---|--------------------|--------------------|------------------|--------------------|--------------------|------------------|
| Dependent variable | All students | Female | Male | Black | White | Other races |
| | GPA | | | | | |
| Average treatment effect of tracking on the treated | -.123*** (.031) | -.116*** (.039) | -.122** (.05) | -.232*** (.086) | -.138*** (.043) | .141** (.069) |
| Dormitory fixed effects | × | × | × | × | × | × |
| Year fixed effects | × | × | × | × | × | × |
| Individual controls | × | × | × | × | × | × |
| # dorm-year clusters | 60 | 36 | 42 | 60 | 58 | 59 |
| # dormitory students | 6600 | 3368 | 3232 | 3234 | 2428 | 845 |
| # other students | 6685 | 3459 | 3226 | 689 | 3484 | 2376 |

Notes: Standard errors in parentheses are estimated using a cluster-robust variance estimator that allows unrestricted error correlation at the dormitory-year level, with non-dormitory students treated as singleton clusters. ***, ** and * denote significance at 1%, 5% and 10% levels respectively. Observations with missing individual controls are excluded from the sample. All columns control for student gender, race, language, nationality, a cubic in high school graduation test scores and all possible three-way interactions.

Table 1.5: Estimates of the average and inequality treatment effects of tracking using a non-linear difference-in-differences model with and without reweighting to adjust for differences in baseline student characteristics.

| Sample | (1) | (2) | (3) | (4) | (5) | (6) |
|--------------------------|---------|--------------------------------|-------------------|---------|-------------------------------------|-------------------|
| | Treated | All students Counterfactual | Difference | Treated | Restricted sample Counterfactual | Difference |
| Mean | .052 | .182 | -.131* (.072) | .084 | .178 | -.095* (.056) |
| Interdecile range | 2.238 | 1.877 | .361*** (.106) | 2.185 | 1.910 | .275*** (.102) |
| Interquartile range | 1.023 | .911 | .112** (.056) | 1.016 | .916 | .100* (.056) |
| Gini coefficient | .097 | .079 | .018*** (.004) | .094 | .080 | .014*** (.004) |
| Coefficient of variation | 1.016 | 1.010 | .006*** (.001) | 1.015 | 1.010 | .005*** (.001) |
| Individual controls | | | | × | × | × |
| Dormitory fixed effects | | | | × | × | × |

Notes: Standard errors in parentheses are estimated from 1000 bootstrap replications, stratifying by tracking/random assignment period and by dormitory/ non-dormitory status, clustering at the dormitory-year level, and treating non-dormitory students as singleton clusters. ***, ** and * denote significance at 1%, 5% and 10% levels respectively. Observations with missing individual controls are excluded from the restricted sample. Columns (4), (5), and (6) use propensity score reweighting to control for dormitory fixed effects and student gender, race, language, nationality, a cubic in high school graduation test scores and all possible three-way interactions.

Table 1.6: Transition probabilities for students from their rank in the distribution of high school graduation test scores to their rank in the distribution of university GPA

| | | Panel A: Tracking | | | | Panel B: Randomization | | | |
|--------------|---|-----------------------------|-----|-----|-----|-----------------------------|-----|-----|-----|
| | | Quartiles of university GPA | | | | Quartiles of university GPA | | | |
| | | 4 | 3 | 2 | 1 | 4 | 3 | 2 | 1 |
| Quartiles of | 4 | .37 | .22 | .13 | .28 | .39 | .25 | .15 | .21 |
| high school | 3 | .29 | .25 | .19 | .27 | .28 | .23 | .23 | .26 |
| graduation | 2 | .13 | .18 | .28 | .41 | .16 | .19 | .27 | .39 |
| test scores | 1 | .05 | .04 | .23 | .68 | .06 | .08 | .22 | .63 |

Notes: Quartiles are defined on the full sample of all dormitory and non-dormitory students in each period but transition probabilities are reported for dormitory students only.

Table 1.7: Treatment effects of tracking on students' academic mobility, measured by the probability of changing their rank in the distribution of high school graduation test scores to their rank in the distribution of university GPA

| | Tracking | Randomization | Difference |
|--|----------|---------------|-----------------|
| Average probability of changing quartile | .813 | .830 | -.017 (.021) |
| Average number of quartiles changed | .921 | .912 | .009 (.038) |
| “Correlation” between high school and college quartile | .547 | .566 | -.019 (.035) |

Notes: Quartiles are defined on the full sample of all dormitory and non-dormitory students in each period but transition probabilities are reported for dormitory students only. Standard errors in parentheses are estimated from 1000 bootstrap replications, stratifying by tracking/random assignment period and by dormitory/ non-dormitory status, clustering at the dormitory-year level, and treating non-dormitory students as singleton clusters. ***, ** and * denote significance at 1%, 5% and 10% levels respectively.

Table 1.8: Estimates of the effect of peers' high school graduation test scores on students' university GPAs using a linear-in-means model for peer effects

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|--------------------------------------|-------------------|-------------------|-------------------|-------------------|-------------------|
| Sample | Randomly assigned dormitory students | | | | | |
| Dependent variable | GPA | | | | | |
| Own high school graduation test score | .362*** (.017) | .332*** (.016) | .331*** (.016) | .363*** (.017) | .333*** (.016) | .332*** (.016) |
| Mean dorm high school graduation test score | .241** (.089) | .222** (.092) | .22** (.098) | .35*** (.075) | .336*** (.067) | .216** (.099) |
| Demographic controls | | × | × | | × | × |
| Dormitory fixed effects | | | × | | | × |
| Excluding outlying low-SES dormitory | | | | × | × | × |
| Adjusted R^2 | .213 | .236 | .248 | .215 | .239 | .248 |
| # observations | 3068 | 3068 | 3068 | 3048 | 3048 | 3048 |
| # clusters | 30 | 30 | 30 | 28 | 28 | 28 |

Notes: Standard errors in parentheses are estimated using a cluster-robust variance estimator that allows unrestricted error correlation at the dormitory-year level, with non-dormitory students treated as singleton clusters. ***, ** and * denote significance at 1%, 5% and 10% levels respectively. Observations with missing high school graduation test scores are excluded from the sample. Columns (2), (3), (5), and (6) control for students' gender language, nationality, and race.

Table 1.9: Estimates of the effect of peers' high school graduation test scores on students' university GPAs using a quadratic-in-means model for peer effects

| | | | |
|--|-------------------|-------------------|-------------------|
| Own high school graduation test score | .4*** (.023) | .374*** (.022) | .373*** (.022) |
| Own high school graduation test score squared | .137*** (.017) | .143*** (.018) | .142*** (.018) |
| Mean dorm high school graduation test score | .221*** (.058) | .174** (.08) | .316*** (.145) |
| Mean dorm high school graduation test score squared | .306*** (.089) | .273*** (.096) | -.159 (.178) |
| Own \times mean dorm high school graduation test score | -.129** (.063) | -.132** (.06) | -.132** (.06) |
| p -value of test against linear-in-means model | 0 | 0 | 0 |
| Demographic controls | | × | × |
| Dormitory controls | | × | |
| Dormitory fixed effects | | | × |
| Adjusted R^2 | .244 | .272 | .278 |
| # observations | 3068 | 3068 | 3068 |
| # clusters | 30 | 30 | 30 |

Notes: Standard errors in parentheses are estimated using a cluster-robust variance estimator that allows unrestricted error correlation at the dormitory-year level, with non-dormitory students treated as singleton clusters. ***, ** and * denote significance at 1%, 5% and 10% levels respectively. Observations with missing high school graduation test scores are excluded from the sample. Columns (2) and (3) control for students' gender language, nationality, and race. Column (2) controls for dormitories' time-invariant physical characteristics: proximity to the university's main academic campus, number of students, and whether students have single or shared rooms.

Table 1.10: Estimates of the effect of peers' high school graduation test scores on students' university GPAs, separately for peers in the students' own and different race groups

| Sample Dependent variable | (1) | (2) | (3) | (4) |
|---|--|-------------------|-------------------|-------------------|
| | Randomly assigned dorm students GPA | | | |
| Own high school (HS) graduation test score | .362*** (.017) | .331*** (.016) | .327*** (.017) | .327*** (.017) |
| Mean dorm HS graduation test score | .241** (.089) | .22** (.098) | | |
| Mean dorm HS graduation test score for own race group | | | .204*** (.058) | .159* (.09) |
| Mean dorm HS graduation test score for other race groups | | | -.003 (.054) | -.051 (.066) |
| <i>p</i> -value for test of equal effects within and across races | | | 0 | .04 |
| Demographic controls | | × | | × |
| Dormitory fixed effects | | × | | × |
| Adjusted R ² | .213 | .248 | .22 | .248 |
| # observations | 3068 | 3068 | 3068 | 3068 |
| # clusters | 30 | 30 | 30 | 30 |

Notes: Standard errors in parentheses are estimated using a cluster-robust variance estimator that allows unrestricted error correlation at the dormitory-year level, with non-dormitory students treated as singleton clusters. ***, ** and * denote significance at 1%, 5% and 10% levels respectively. Observations with missing high school graduation test scores are excluded from the sample. Columns (2) and (4) control for students' gender, language, nationality, and race.

Table 1.11: Estimates of the effect of peers' high school graduation test scores on students' university GPAs, separately for peers in the students' own and different college/faculty/school (e.g. commerce, engineering, social science)

| Sample Dependent variable | (1) | (2) | (3) |
|---|--|-------------------|-------------------|
| | Randomly assigned dorm students GPA | | |
| Own high school graduation test score | .345*** (.019) | .298*** (.017) | .307*** (.016) |
| Mean dorm high school graduation test score for own faculty | .049 (.047) | .105** (.044) | .066 (.062) |
| Mean dorm high school graduation test score for other faculties | .152* (.072) | .171** (.062) | .151 (.073) |
| <i>p</i> -value for test of equal effects within and across faculties | .171 | .342 | .409 |
| Demographic controls | | × | × |
| Dormitory fixed effects | | × | × |
| Faculty fixed effects | | | × |
| Adjusted R ² | .202 | .239 | .251 |
| # observations | 3068 | 3068 | 3068 |
| # clusters | 30 | 30 | 30 |

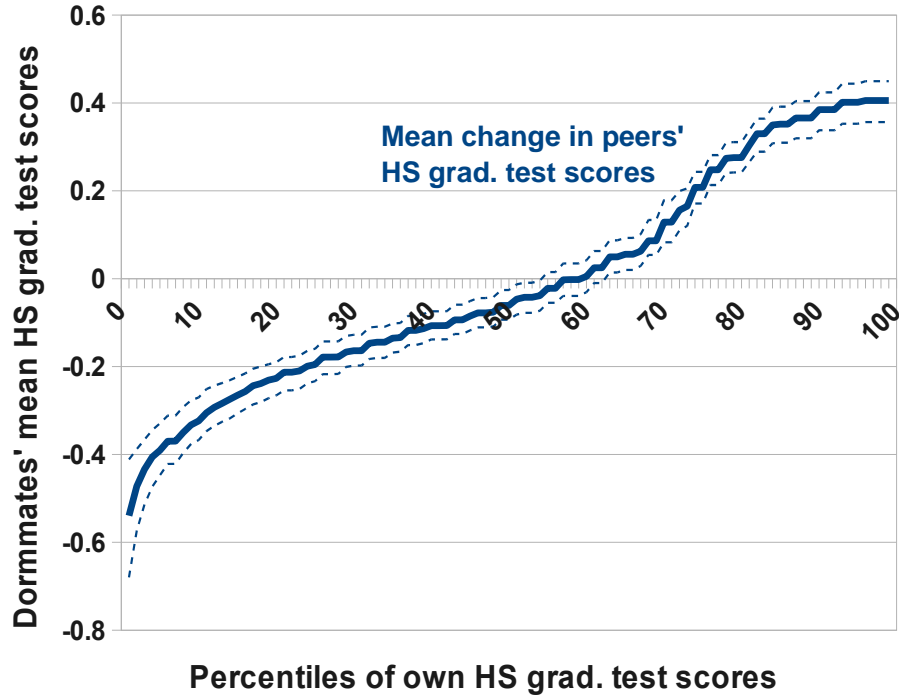
Notes: Standard errors in parentheses are estimated using a cluster-robust variance estimator that allows unrestricted error correlation at the dormitory-year level, with non-dormitory students treated as singleton clusters. ***, ** and * denote significance at 1%, 5% and 10% levels respectively. Observations with missing high school graduation test scores are excluded from the sample. Columns (2) and (3) control for students' gender, language, nationality, and race.

Table 1.12: Average treatment effects of tracking on different measures of university academic performance estimated using a linear difference-in-differences regression model

| Sample | (1) All | (2) Restricted | (3) All | (4) Restricted | (5) All | (6) Restricted |
|---|------------------|--------------------|--|-------------------|---|-------------------|
| Dependent variable | GPA | | Mean GPA from courses where students are not academically excluded | | Prop. of courses where students are academically excluded | |
| Average treatment effect of tracking on the treated | -.129* (.073) | -.123*** (.031) | -.076 (.089) | -.068** (.033) | .028*** (.006) | .027*** (.005) |
| Dep var mean | | | | | .05 | .049 |
| Dormitory fixed effects | | × | | × | | × |
| Year fixed effects | | × | | × | | × |
| Individual controls | | × | | × | | × |
| # dorm-year clusters | 60 | 60 | 60 | 60 | 60 | 60 |
| # dormitory students | 7480 | 6600 | 7449 | 6576 | 7480 | 6600 |
| # other students | 7188 | 6685 | 7043 | 6559 | 7188 | 6685 |

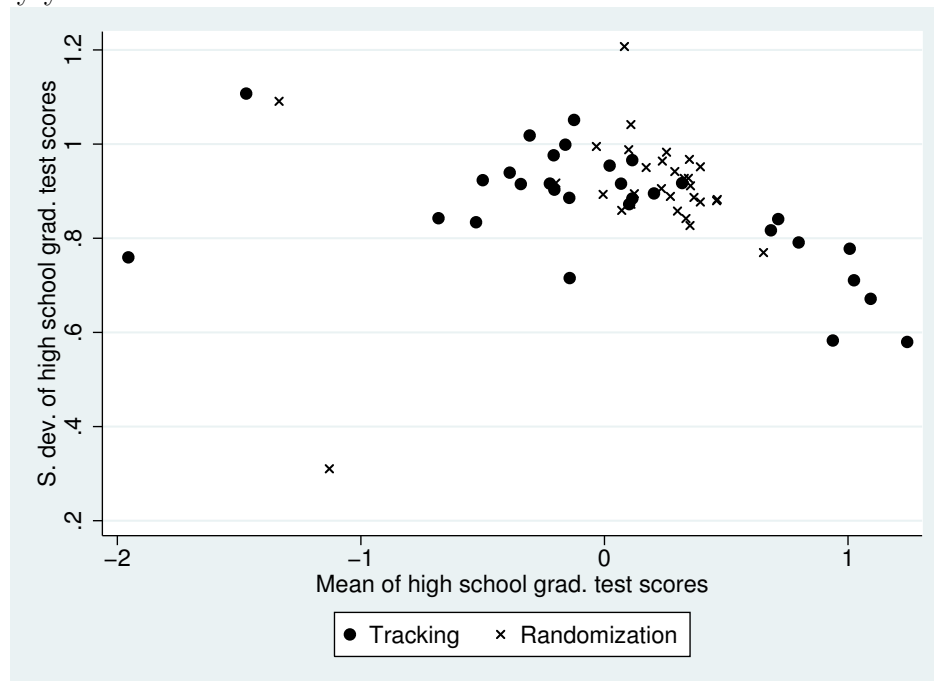
Notes: Standard errors in parentheses are estimated using a cluster-robust variance estimator that allows unrestricted error correlation at the dormitory-year level, with non-dormitory students treated as singleton clusters. ***, ** and * denote significance at 1%, 5% and 10% levels respectively. Observations with missing individual controls are excluded from the sample. Columns (2), (4), and (6) control for student gender, race, language, nationality, a cubic in high school graduation test scores and all possible three-way interactions. Columns (1) and (2) calculate students' GPA using all courses for which they register and assigning zero grades if they are academically excluded from the course. Columns (3) and (4) calculate students' GPA using only courses from which they are not excluded. Columns (5) and (6) measure the percentage of courses from which students were excluded.

Figure 1.1: Difference between tracking and randomization periods in the mean high school graduation test scores of students' dormitory peer groups for students at each percentile of the high school graduation test score distribution.



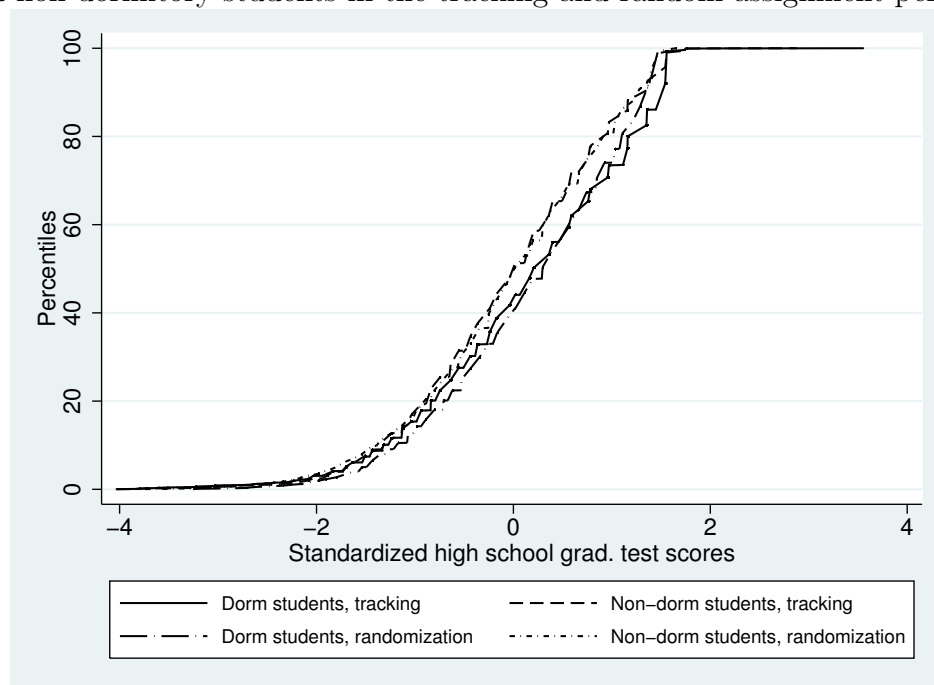
Notes: This figure quantifies the change in students' mean peer group academic proficiency, measured by their high school graduation test scores. It shows that low-scoring students have significantly lower-scoring dormitory peers in the tracking period (≈ 0.5 standard deviations in the lowest decile) and high-scoring students have significantly higher-scoring dormitory peers (≈ 0.4 standard deviations in the top decile). The figure is constructed in four steps. First, I calculate the mean high school graduation test score in each dormitory and assign this to each student as a measure of their peers' academic proficiency. Second, I estimate a nonparametric regression of this peer group measure against students' own high school graduation test scores, separately for the tracking and random assignment periods. I use a local linear regression with an Epanechnikov kernel and a plug-in bandwidth following Fan and Gijbels (1996). I allow the bandwidth to differ for the tracking and random assignment periods. Third, I calculate the difference between the two fitted curves at each percentile of the distribution of high school graduation test scores. Finally, I construct a 95% confidence interval from the 2.5 and 97.5 percentiles of a nonparametric bootstrap, stratifying by tracking/random assignment period.

Figure 1.2: Mean and standard deviation of high school graduation test scores in each dormitory-by-year



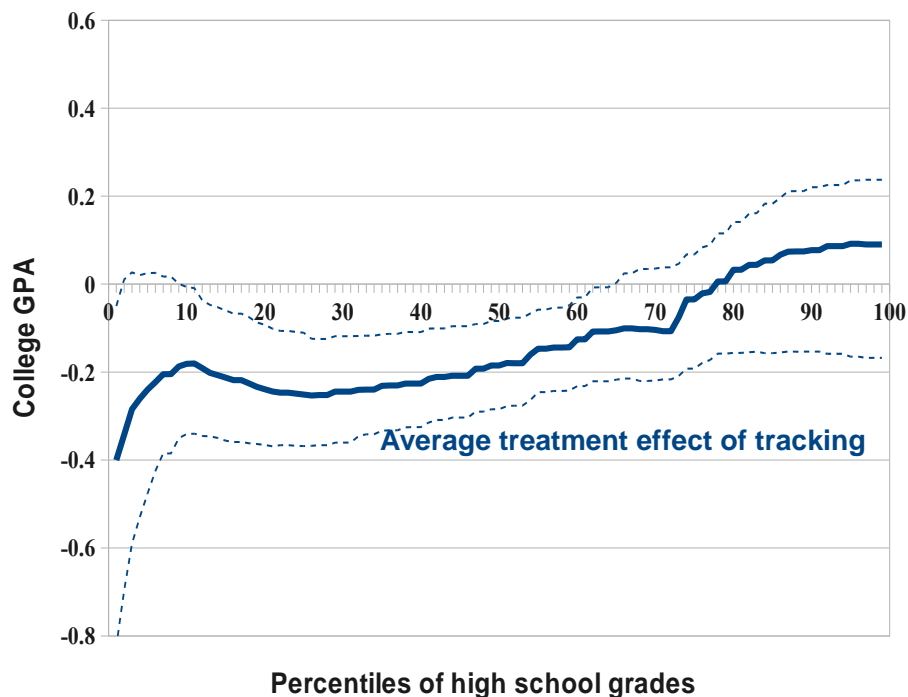
Notes: This figure shows the distribution of dormitory-by-year means and standard deviations of high school graduation test scores. The tracking policy increases the observed range of mean dormitory test scores to $(-0.2, 0.6)$ from $(-0.8, 1.2)$ under the random assignment policy. This provides a measure of the redistribution of peer groups induced by the policy change. The four outliers in the left tail are the four annual observations of a single dormitory with a different assignment rule that typically houses only students from low-income families. It accounts for $\approx 1.4\%$ of the sample and all results are robust to its exclusion.

Figure 1.3: Cumulative distribution function of high school graduation test scores for dormitory and non-dormitory students in the tracking and random assignment periods



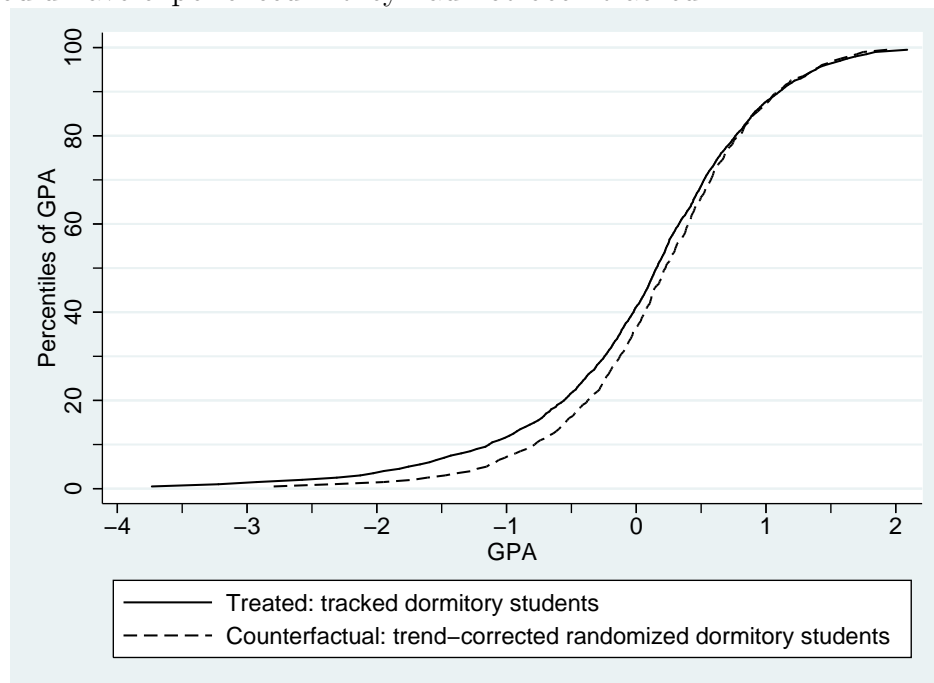
Notes: This figure shows the distribution of standardized high school graduation test scores. The long left tail increases the within-dormitory variation of this measure in low-track dormitories.

Figure 1.4: Average treatment effects of tracking for students at each percentile of the distribution of high school graduation test scores



Notes: Point estimates are obtained by estimating a local linear regression of GPA on high school graduation test scores, separately for dormitory students in the tracking period, dormitory students in the random assignment period, non-dormitory students in the tracking period, and non-dormitory students in the random assignment period; I then take the second difference between these estimates to obtain local difference-in-differences estimates. The local linear regressions use Epanechnikov kernels and plug-in bandwidths chosen separately for each of the four nonparametric regressions following Fan and Gijbels (1996). The results are qualitatively robust to substantial changes in the bandwidth parameter. Confidence intervals are estimated using a percentile bootstrap with 1000 replications, stratifying by tracking/random assignment period and by dormitory/non-dormitory status, clustering at the dormitory-year level, and treating non-dormitory students as singleton clusters. Observations with missing high school graduation test scores are excluded from the sample.

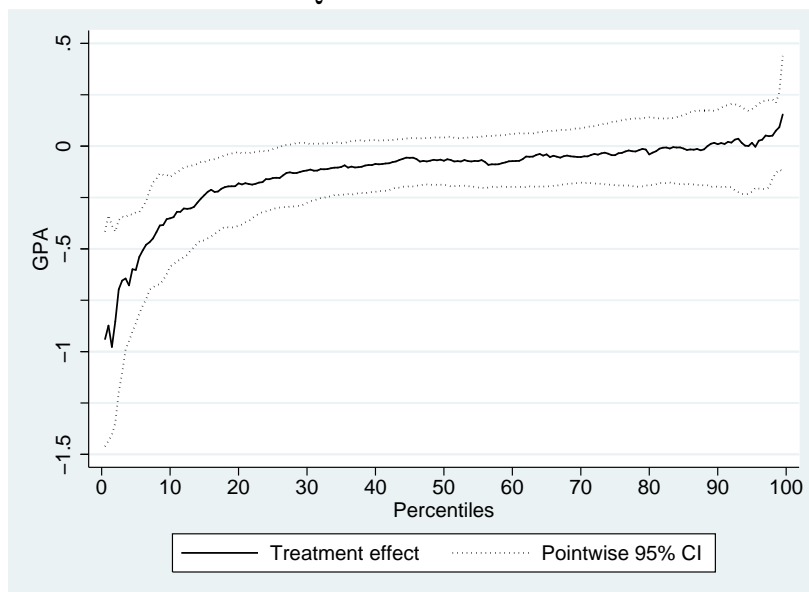
Figure 1.5: Nonlinear difference-in-differences analysis, comparing the observed distribution of GPA for tracked dormitory students to the counterfactual distribution that the same students would have experienced if they had not been tracked



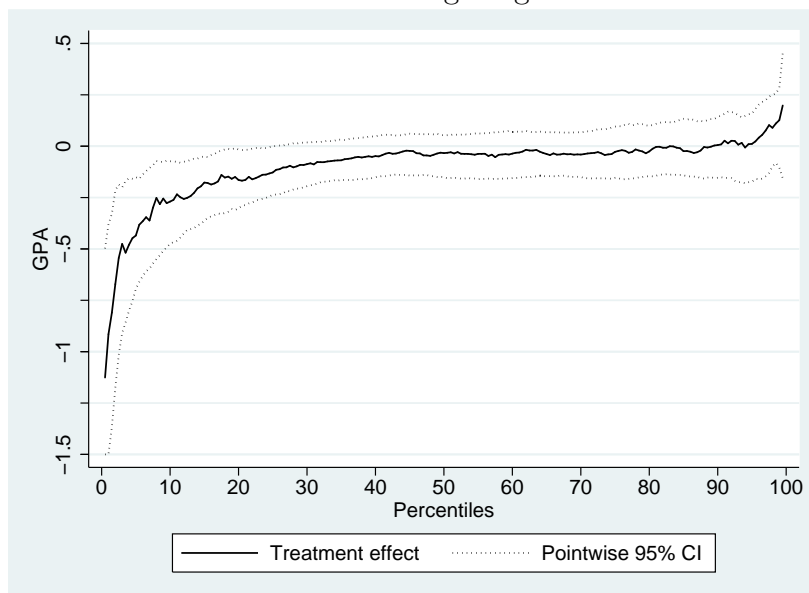
Notes: The counterfactual cumulative distribution function is estimated using the nonlinear difference-in-differences model discussed in appendix 1.B, without adjusting for differences in the distribution of baseline covariates between the tracking and randomization periods.

Figure 1.6: Nonlinear difference-in-differences analysis, which uses the observed and counterfactual distributions of GPA for tracked dormitory students to estimate quantile treatment effects of tracking

Panel A: Quantile treatment effects

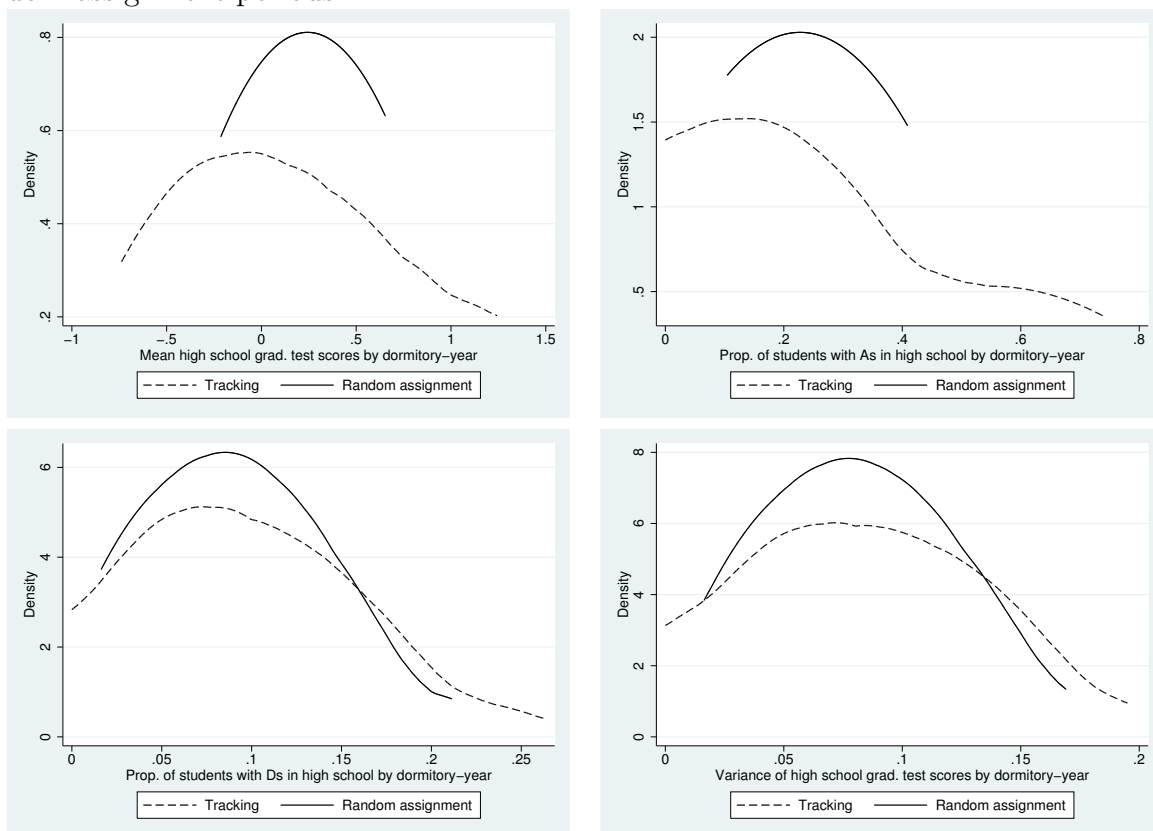


Panel B: Quantile treatment effects after reweighting for observed student characteristics



Notes: The counterfactual distribution of GPAs is estimated using the nonlinear difference-in-differences model discussed in appendix 1.B. Panel B adjusts for differences in the distribution of observed student characteristics between the tracking and randomization periods using reweighting. The reweighting model includes gender, language, nationality, race, a cubic in high school graduation test scores, all possible three-way interactions, and dormitory fixed effects. The confidence intervals are estimated from 1000 replications of a percentile bootstrap, clustering at the dormitory-year level and stratifying by tracking/randomization period and dormitory/non-dormitory group.

Figure 1.7: Density of selected summary statistics for high school graduation test scores at the dormitory-by-year level, showing how these densities differ between the tracking and random assignment periods



Notes: These comparative densities show that there are values of the dormitory-level summary statistics that are observed under tracking but not under random assignment, such as mean high school graduation test scores below -0.2 and above 0.6. This emphasizes the extent to which using peer effects estimated under random assignment to predict the effects of tracking requires out-of-sample extrapolation. The density plots are deliberately oversmoothed (i.e. artificially high values of the bandwidth are used) to focus attention on the range.

1.A Robustness Checks

1.A.1 *Changes in course taking behavior*

The body of the paper treated GPAs during the tracking and random assignment periods as comparable measures, an assumption that may be problematic if the mix of courses taken is different during the two periods and grading standards vary across courses. I perform three robustness checks to show that the average treatment effects discussed in section 1.3 are not driven by changes in the mix of courses. In all of these checks, I continue to condition on observed student characteristics and dormitory fixed effects as in the body of the paper. First, column 2 of table 1.13 shows that the estimated treatment effect is largely unaffected by the inclusion of college/faculty/school fixed effects. Second, columns 3 to 8 report treatment effects estimated separately for each of the university's six colleges/faculties/schools: commerce, engineering, law, medicine, science, and social science and humanities. The point estimates are all negative, four of the six are significantly different to zero, and only one is significantly different to the overall treatment effect. The one exception is the law faculty, which is sufficiently small that the extremely large and imprecisely estimated negative point estimate should be treated with caution. Third, table 1.14 estimates treatment effects within the six largest introductory courses at the university: economics, information systems, management, physics for engineers, sociology, and statistics. All treatment effects are negative, four of the six are significantly different to zero, and none of the six are significantly different to the overall treatment effect. These results show that the treatment effect is not driven by changes in course selections but instead occurs within individual courses.

1.A.2 *Instrumental variables estimates based on admission rules*

The identification strategy in the bulk of the paper can be characterized as a selection on observed variables strategy. I claim that after conditioning on students' demographic characteristics and high school graduation test scores, the change in GPAs between the tracking and random assignment periods would be identical for dormitory and non-dormitory students if tracking were not in place. If students choose whether to live in dormitories or in private accommodation or if students choose whether to attend the university during

the tracking or random assignment period in order to maximize their expected GPA, this assumption may be violated.

I argue in section 1.2 that such selection on unobserved characteristics is implausible, given the limited information available to prospective students about the dormitory allocation policies and the rules that determine which students are admitted to the dormitory system. In this appendix, I explicitly use these admission rules to construct instrumental variables that exogenously affect the probability that students will be in each of the four groups in my analysis (dormitory and non-dormitory students, during the tracking and random assignment periods).

Specifically, I use the year of high school graduation as an instrument for the period that students enter the university: the instrument takes the value one for students who graduated in 2004 or earlier and the tracking indicator equals one for students who enrolled in the university in 2005 or earlier.³⁴ I use the location of each student's high school as an instrument for whether the student enters the dormitory system: the instrument takes the value one for students who attended a high school outside Cape Town and the dormitory indicator equals one for students living in a dormitory. I use the interaction of these instruments as an instrument for the treatment indicator, which equals the interaction of dormitory and the tracking indicators.

The instruments are strongly correlated with the group indicators: the first stage coefficients for the tracking instrument, dormitory instrument, and interaction are 0.927 (standard error 0.006), 0.647 (0.014), and 0.706 (0.031) respectively. The exclusion restriction for the instrumental variables difference-in-differences model is that any direct effect of the year of high school graduation on GPA (*i.e.* any effect not operating through the dormitory assignment policy) is identical for dormitory and non-dormitory students and that any direct effect of whether the student attended a high school inside or outside Cape Town does not change from the tracking to the mixing period. This restriction does not rule out time trends in student characteristics or differences in school quality but requires that these are, respectively, equal for high schools inside and outside Cape Town and unchanging through time.

³⁴South Africa uses a January to December school year, so this means that students who graduate in December 2004 or earlier are defined as eligible for the tracking period, which included the academic year starting in January 2005.

Table 1.15 shows that the instrumented treatment effects are similar to those estimated without the instrumental variables correction. They are, however, relatively imprecisely estimated. This may in part reflect a limitation in the “high school in or outside Cape Town” instrument. Students are admitted to the dormitory system based on their home address and the address of the high school that they attend is an imperfect proxy, as some students attend schools far from their home and live in school hostels or dormitories.³⁵ This means that the instrument suffers from non-classical measurement error. Simulation results suggest that measurement error-induced bias increases the absolute value of the estimated local average treatment effect.

Given this concern, caution should be exercised in interpreting the instrumental variables estimates. They provide reassuring evidence that the treatment effects reported in the body of the paper are not driven by selection on unobserved variables but should not be viewed as a definitive refutation of this possibility. Appendix 1.A.4 therefore develops a further sensitivity analysis that demonstrates the robustness of the least squares treatment estimates to at least some forms of selection on unobserved variables.

1.A.3 Time trends in student characteristics

The identifying assumption of the difference-in-differences model will be violated if the trends in dormitory and non-dormitory students’ outcomes differed for reasons unrelated to the policy change. I use a simple placebo test to explore this possibility. Specifically, I compare the change in dormitory students’ mean GPA from 2001–2002 to 2004–2005 to the change in non-dormitory students’ GPA over the same time period.³⁶ The tracking policy was in place during this entire period, so under the identifying assumption this estimate should be approximately zero. The first column of table 1.16 shows that the difference in these pre-treatment trends was -0.04 standard deviations of GPA (standard error 0.04). This estimate is insignificant and considerably smaller than the treatment effect using the actual experiment. This supports the plausibility of the identifying assumption. The second column

³⁵I do not observe the home address for the majority of students and so cannot use this information directly.

³⁶I omit 2003 from the analysis because a database problem at the University of Cape Town resulted in over half the observations in 2003 being assigned non-unique identifiers.

performs the same test for students' mean high school graduation test scores. This point estimate is also small and insignificant (-0.02, standard error 0.04), suggesting that the pre-treatment trends in the characteristics of students entering the university were also equal and again supporting the plausibility of the identifying assumption.³⁷

1.A.4 *Sensitivity analysis for potential violations of the identifying assumptions*

The difference-in-differences model used in the bulk of the paper yields consistent estimators of the average treatment effects of tracking on the treated under specific assumptions. In particular, the linear model assumes that the mean change in dormitory students' unobserved characteristics between the two periods was equal to the mean change in non-dormitory students' unobserved characteristics. This assumption is fundamentally untestable but the discussion and data presented in section 1.2 suggest that it is reasonable in this context. Furthermore, the fact that the estimated treatment effects are highly robust to controls for observed student characteristics suggest that there are no substantial differences in unobserved characteristics that are correlated with the observed characteristics.

However, there may still be concerns that there exist some unobserved characteristics that satisfy three conditions: they are correlated with GPA, their distribution violates the "equal trends" assumption above, and they are weakly correlated or uncorrelated with the observed characteristics. Here I propose a simple model of this form of selection and show what it implies for the estimated value of the treatment effect. Consider the possibility that GPAs are generated by the model

$$\begin{aligned} GPA &= \delta_{11}DT + \delta_{10}D(1 - T) + \delta_{01}(1 - D)T + \delta_{00}(1 - D)(1 - T) + Z\gamma + \epsilon \\ &\equiv \delta G + Z\gamma + \epsilon \end{aligned} \tag{1.9}$$

where $GPA_{n \times 1}$ is a vector of outcomes, $G_{n \times 4}$ is a matrix indicating group membership (dormitory in tracking period, dormitory in random assignment period, non-dormitory in

³⁷I do not have access to dormitory assignment records for 2001 and 2002. The placebo test results are therefore intention-to-treat estimates using high school location in the regression instead of dormitory/non-dormitory status. This also means that I cannot define the dormitory-by-year clusters I would normally use in the variance estimation. I therefore report heteroscedasticity-robust standard errors in column (1), which produce an artificially tight confidence interval. As I find an insignificant and imprecisely estimated effect, this should not be a source of concern.

tracking period, non-dormitory in random assignment period), Z is an unobserved student characteristic, and $\epsilon_{n \times 1}$ is a mean-zero error term uncorrelated with GPA , G , and Z . To reduce the dimension of the problem, I assume that $Z \in \{-1, 1\}$ with $\mathbb{E}[Z] = 0$, $\mathbb{E}[Z|D = 0, T = 1] = \mathbb{E}[Z|D = 0, T = 0] = 0$,³⁸ and $\mathbb{E}[Z|D = 1, T = 1] = \rho = -\mathbb{E}[Z|D = 1, T = 0]$. Substantively, I interpret Z as “academic orientation,” and consider two possible forms of selection:

- Students with high academic orientation ($Z = 1$) prefer to live in dormitories under tracking than under random assignment because this exposes them to similar peers who are also focused on academic performance, so $\rho > 0$.
- Students with low academic orientation ($Z = -1$) prefer to live in dormitories under tracking than under random assignment because this exposes them to similar peers who are also more interested in social and leisure activities, so $\rho < 0$.

This might arise because the set of students who apply to the university differs in their values of Z between the two periods, not necessarily because the same students are being redistributed between the two periods.

As Z is unobserved, the misspecified GPA model is $GPA = \delta G + \epsilon$. Estimating this by least squares yields

$$\begin{pmatrix} \hat{\delta}_{11} \\ \hat{\delta}_{10} \\ \hat{\delta}_{01} \\ \hat{\delta}_{00} \end{pmatrix} \longrightarrow_p \begin{pmatrix} \delta_{11} + \gamma \rho \frac{\mu_{11}}{\sigma_{11}^2} \\ \delta_{10} - \gamma \rho \frac{\mu_{11}}{\sigma_{11}^2} \\ \delta_{01} \\ \delta_{00} \end{pmatrix}$$

where μ_{dt} and σ_{dt}^2 are the mean and variance respectively of the indicator variable $\mathbf{1}\{D = d, T = t\}$. The difference-in-differences test statistic is therefore

$$\hat{\tau} = \hat{\delta}_{11} - \hat{\delta}_{10} - \hat{\delta}_{01} + \hat{\delta}_{00} \longrightarrow_p \tau + 2\gamma\rho \left(\frac{\mu_{11}}{\sigma_{11}^2} + \frac{\mu_{10}}{\sigma_{10}^2} \right). \quad (1.10)$$

³⁸The assumption that the distribution of Z is identical for non-dormitory students across the two periods simplifies the resultant algebra but can be relaxed without altering the conclusions.

The test statistic is upward biased when academically orientated students are more common in the tracking period ($\rho > 0$) and downward biased when they are less common ($\rho < 0$). Furthermore, the means and variances can be replaced by sample analogues and a “bias-corrected” estimator³⁹

$$\hat{\tau}^{BC} = \hat{\tau} - 2\gamma\rho \left(\frac{\hat{\mu}_{11}}{\hat{\sigma}_{11}^2} + \frac{\hat{\mu}_{10}}{\hat{\sigma}_{10}^2} \right). \quad (1.11)$$

can be constructed for any hypothesized values of ρ and γ .

I use the observed covariates to calibrate plausible values for these parameters. In particular, I note that if the observed binary characteristics in table 1.1 are transformed so that $X \in \{-1, 1\}$, the largest value of $|\rho|$ is 0.095 (for language). Similarly, the value of γ can be chosen to match the strength of the association between *GPA* and selected student demographic characteristics. For example, the value of γ implied by the difference between black and white GPAs is 0.23.

Figure 1.8 plots the value of the bias-corrected treatment effect for selected values of γ and for all $-0.2 \leq \rho \leq 0.2$. The top panel shows that if the unobserved characteristic differs between the tracking and random assignment periods as much as the “most different” observed characteristic (language) and is “as important” a determinant of GPA as race, the bias-corrected treatment effect is approximately -0.07 standard deviations (if academically orientated students select out of tracking) or -0.17 standard deviations (if academically orientated students select into tracking).⁴⁰

The second panel repeats the same analysis under the assumption that the true GPA model is $GPA = \delta G + X\beta + Z\gamma + \epsilon$, the misspecified model is $GPA = \delta G + X\beta + \epsilon$, and $\mathbb{E}[Z|X] = 0$. The last assumption is not necessary but it again simplifies the algebra and corresponds to the “worst-case” selection problem, as the bias arising from an unobserved Z that is correlated with X will be reduced by controlling for X . I include in X students’ gender, language, nationality, and race, a cubic in high school graduation test scores, all possible threeway interactions and dormitory fixed effects.⁴¹ Conditioning on X marginally

³⁹This estimator is actually still biased for τ because $\hat{\mu}_{dt}/\hat{\sigma}_{dt}^2$ is a consistent but biased estimator of μ_{dt}/σ_{dt}^2 . The bias arising from the nonlinearity of the statistic can be reduced using a higher order Taylor series approximation. This correction makes no difference to my results so I omit it for expositional clarity.

⁴⁰Inference on these point estimates should be performed with a degree of caution, as the omission of Z from the GPA model will also inflate the estimated standard errors of the coefficients.

⁴¹The previous bias calculation can be readily extended to control for X using standard partitioned

attenuates the bias-corrected treatment effects but the difference is negligible, which is consistent with the limited impact on the treatment effect of controlling for individual and dormitory characteristics throughout the paper.

This analysis suggests that an implausibly large degree of selection on unobserved characteristics is necessary to explain the estimated treatment effect. This, combined with the results of the previous appendix, strongly suggests that the treatment effect is not driven by violations of the identifying assumptions laid out in section 1.2.

1.A.5 *Bootstrap-based test procedures*

My data is characterized by a natural group structure, as is almost inevitable in any study of peer effects. Dormitory students live together and there may be concerns that the error terms are correlated within these buildings. The main body of the paper addresses this problem using a “cluster-robust variance estimator” that generalizes the Eicker-White heteroscedasticity-robust variance estimator. This estimator has the desirable feature of imposing no structure on the nature of the error correlations within dormitory-year clusters⁴² but it converges with the square of the number of clusters, rather than the number of students. Bertrand, Duflo, and Mullainathan (2004) note that this may lead to tests with sizes far lower than their nominal rates when the number of clusters is small. Cameron, Miller, and Gelbach (2008) discuss several remedies for this problem and particularly recommend tests based on either cluster or wild cluster bootstraps applied to pivotal statistics.⁴³

Table 1.17 repeats the analyses in table 1.2, but implements the test that each treatment effect equals zero using four different strategies: the default cluster-robust variance estimator, a cluster bootstrap estimate of the treatment effect’s standard error, a cluster bootstrap

regression results. The only substantive change is that γ is now interpreted as the relationship between GPA and Z conditional on X and $\hat{\tau}$ is obtained from a regression of $M_X(GPA)$ on M_XG , where $M_X = I - X(X'X)^{-1}X'$.

⁴²In contrast, both random effects estimators and Moulton-style corrections to ordinary least squares estimators assume that the group and individual components of the error structure are additively separable. My inferences are robust to using both of these estimators but, as there is no theory-driven reason to assume additive separability, I do not report these results.

⁴³Pivotal statistics are those whose asymptotic distribution does not depend on unknown parameters. This includes most conventional test statistics, as test statistics for single and multiple hypotheses are asymptotically $\mathcal{N}(0, 1)$ and $\chi^2(k)$ respectively under standard assumptions. However, this excludes most parameter and standard error estimators, as these are centered around their true but unknown values.

approximation of the distribution of the test statistic, and a wild cluster bootstrap approximation of the distribution of the test statistic.⁴⁴ Panel A shows that the p -values on the test of zero treatment effect are highly robust across different testing procedures. In most cases, the models that do not control for covariates generate estimates that are marginally significant ($0.014 \leq p \leq 0.187$), while those that control for covariates generate highly significant estimates ($0 \leq p \leq 0.007$). Given this robustness, the use of analytical standard errors as a default is relatively innocuous. Furthermore, the third and fourth testing strategies apply only to test statistics and do not yield valid standard error estimates, limiting the ability of the reader to perform tests other than those reported by the author or to mentally construct confidence intervals.

Panel B of table 1.17 reports the results of implementing the same procedures but allowing for error correlation at the *dormitory level*, rather than *dormitory-year level*. Bertrand, Duflo, and Mullainathan (2004) recommend this approach for most difference-in-differences designs that rely on repeated cross-sections as cluster-specific shocks may persist through time. This concern is arguably less relevant in my application because my sample consists of only first year students, so no student appears in the sample in multiple years. This consideration is relevant only to the extent that a shock affecting first year students in dormitory d in year t continues to affect these students in year $t + 1$ and hence affects the new cohort of first year students in dormitory d . My inferences are again robust to this more conservative inference procedure: the models that control for covariates still yield highly significant estimates.

1.B Reweighted Nonlinear Difference-in-differences Model

Given the limited use of the Athey-Imbens nonlinear difference-in-differences model in the applied literature, this appendix provides an overview of the model and the reweighting extension that I propose.

Begin by defining T as an indicator variable equal to one in the tracking period and zero in the mixing period and D as an indicator variable equal to one for dormitory students and

⁴⁴See Cameron, Miller, and Gelbach (2008) for details on the implementation of these various procedures. I use 1000 replications of a stratified cluster bootstrap that resamples dormitory-year clusters (with probability proportional to size) and individual non-dormitory students after stratifying by period. I implement the bootstrap t -statistic procedures with the relevant null hypothesis imposed.

zero for other students. Formally, the model is identified under three assumptions:

- (A1) GPA in the absence of tracking is strictly continuous and generated by the model $GPA = h(U, T)$, which is monotone in the unobserved scalar U . Note that the function h need not be known and that GPA does not directly depend on D .
- (A2) The distribution of the unobserved characteristic remains constant through time for each group, in this case dormitory and non-dormitory students: $U \perp T | D$.
- (A3) The support of dormitory students' GPAs is contained in the support of non-dormitory students' GPAs: $supp(GPA | D = 1) \subseteq supp(GPA | D = 0)$.

These assumptions are sufficient to identify the counterfactual distribution of dormitory students' GPAs in the tracking period *in the absence of tracking*, denoted by $F_{GPA|D=1,T=1}^{CF}(\cdot)$. These are the outcomes that the treatment group would have experienced in the treatment period if treatment had not been applied. I follow Athey and Imbens in assuming that the distribution of GPA is strictly continuous and has no mass points, so the cumulative distribution function is invertible.⁴⁵

To understand the identification proof, begin by considering the observed distribution of dormitory students' GPAs in the random assignment period, $F_{GPA|D=1,T=0}(\cdot)$. Under assumption A1, GPA can be written as $g = h(U, T)$ and so depends only on the unobservable U and the time period. Under assumption A2, the distribution of U does not change through time, so the difference between $F_{GPA|D=1,T=1}^{CF}(\cdot)$ and $F_{GPA|D=1,T=0}(\cdot)$ is due entirely to the change from $T = 0$ to $T = 1$ in h . If h were known, it would be straightforward to evaluate this change. However, the function is unknown and so an indirect argument must be used to construct a mapping from $g = h(u, 0)$ to $\tilde{g} = h(u, 1)$ for each possible realization u of U .

1. Let A and B denote two students in the random assignment period with GPA g and assume that A is a dormitory student and B is a non-dormitory student. As h does not depend directly on whether students live in a dormitory or not, A and B must have the same value of the unobservable u . The common support assumption A3 ensures

⁴⁵The GPA measure I use is approximately continuous. There are 5215 unique values of GPA, so each value accounts for an average of 0.02% of the total mass. The most common value accounts for only 0.26% of the total mass.

that every dormitory student in the mixing period has an appropriate comparison non-dormitory student with the same GPA.

2. Let $q = F_{GPA|D=0,T=0}(g)$ denote B 's location in the distribution of non-dormitory students in the random assignment period and let C be a non-dormitory student in the tracking period with the same location q in her distribution $F_{GPA|D=0,T=1}(\cdot)$. As h is monotone in u and the distribution of U does not change through time, B and C 's equal locations imply that they have the same value of the unobservable u . Their GPAs differ only due to the change from $T = 0$ to $T = 1$ in $h(\cdot, T)$. Let $\tilde{g} = F_{GPA|D=0,T=1}^{-1}(q) = F_{GPA|D=0,T=1}^{-1}(F_{GPA|D=0,T=0}(g))$ denote C 's GPA.
3. In the absence of tracking, the same analysis could have been applied to dormitory students. Hence, a dormitory student in the tracking period with $U = u$ would have a GPA of $\tilde{g} = F_{GPA|D=0,T=1}^{-1}(q) = F_{GPA|D=0,T=1}^{-1}(F_{GPA|D=0,T=0}(g))$. Denote this student by D .
4. As A and D , the two dormitory students in different periods, have the same values of the unobservable, they would have the same locations in their respective distributions, so $F_{GPA|D=1,T=1}^{CF}(\tilde{g}) = F_{GPA|D=1,T=0}(g)$.

Combining the results in the third and fourth list entries yields the counterfactual distribution of GPAs for dormitory students in the tracking period:

$$F_{GPA|D=1,T=1}^{CF}(\tilde{g}) = F_{GPA|D=1,T=0}\left(F_{GPA|D=0,T=0}^{-1}\left(F_{GPA|D=0,T=1}(\tilde{g})\right)\right) \quad (1.12)$$

The quantile treatment effect on the treated at quantile q is defined as the horizontal distance between the observed and counterfactual distributions:

$$\Delta(q) = F_{GPA|D=1,T=1}^{-1}(q) - F_{GPA|D=1,T=1}^{CF,-1}(q) \quad (1.13)$$

Intuitively, the first assumption $A1$ imposes sufficient structure on the data generating process to allow us to compare students' GPAs across groups and their locations in the GPA distribution through time. The second assumption ensures that changes through time in the

GPA distribution are due to time changes, which are common to both groups, not changes in the distribution of unobserved characteristics.

Note that the model imposes structure only on the data generating process for GPA in the absence of treatment (tracking) and remains agnostic regarding the data generating process for GPA under tracking. This identifies the counterfactual distribution of GPAs for dormitory students under tracking if tracking had not been implemented but provides no information about the counterfactual distribution of GPAs for non-dormitory students if they had been exposed to tracking. The model therefore identifies the *treatment effect on the treated*, in the framework of Heckman and Robb (1985).⁴⁶

The original paper only briefly considers the role of observed student characteristics and suggests two somewhat *ad hoc* means of controlling for these characteristics. The first suggestion is to implement the model separately for specific values of the covariates (for example, for male and female students). However, this is clearly infeasible with many discrete covariates or with continuous covariates. The second suggestion is to “residualize” GPA by regressing it on the covariates and then applying the model to the residuals from this regression. However, this residualization scheme has a number of disadvantages. It assumes that the observed and unobserved characteristics are additively separable in the production function (i.e. $GPA = X\beta + \epsilon$), which undermines the nonparametric spirit of the entire model. Any misspecification of the functional form of X will result in inconsistent estimates of β and hence $\hat{\epsilon}$ may not converge in distribution to $F(\epsilon)$.

I instead use a reweighting scheme that avoids the assumption of additive separability and may be more robust to specification errors. Specifically, I define the reweighted counterfactual distribution as

$$F_{GPA^{11}}^{RW,CF}(g) = F_{GPA_{\omega}^{10}} \left(F_{GPA_{\omega}^{00}}^{-1} (F_{GPA^{01}}(g)) \right) \quad (1.14)$$

⁴⁶Athey and Imbens also consider a “quantile difference-in-differences” model that computes second differences between groups in quantiles of the outcome: $F_{GPA|D=1,T=1}^{-1}(q) - F_{GPA|D=1,T=0}^{-1}(q) - F_{GPA|D=0,T=1}^{-1}(q) - F_{GPA|D=0,T=0}^{-1}(q)$. This is perhaps a more intuitive approach but it requires the stronger assumption that the distribution of the unobservables is identical for both groups in both time periods. This stronger assumption permits a direct comparison of GPAs through time and across groups, rather than using the location comparison through time. This stronger assumption is clearly inappropriate in my application (see table 1.1) but my results are robust to using this alternative model.

where $F_{GPA_{\omega}^{d0}}(\cdot)$ is the distribution function of $GPA \times Pr(T = 1|D = d, X)/Pr(T = 0|D = d, X)$. Intuitively, this scheme assigns high weight to students in the random assignment period whose observed characteristics are similar to those in the tracking period. This is a straightforward extension of the reweighting techniques used in the wage decomposition literature (DiNardo, Fortin, and Lemieux, 1996) and the program evaluation literature (Hirano, Imbens, and Ridder, 2003). Firpo (2007) lays out the technical assumptions under which the reweighted distribution is consistently estimated by the predicted probabilities from a series logistic regression of T on X . Under these assumptions

$$\hat{\tau}^{QTT}(q) = \hat{F}_{GPA^{11}}^{-1}(q) - \hat{F}_{GPA^{11}}^{-1, RW, CF}(q) \quad (1.15)$$

is a consistent estimator of the quantile treatment effect on the treated in the *reweighted nonlinear difference-in-differences* model.

An important assumption invoked for consistency of this reweighted estimator is that the propensity score $Pr(T = 1|D = d, X)$ is consistently estimated. Firpo (2007) proposes doing so using a semiparametric logistic model in which T is regressed on a polynomial function of X whose order satisfies certain regularity conditions. Selecting the order of the polynomial is a difficult process and the literature provides relatively little guidance. In practice, I use polynomial orders from 1 to 4, and the choice of this tuning parameter makes little difference to my results.

I implement the estimator in three steps:

1. I regress an indicator for the tracking period on a flexible logistic function of X , separately for each group, and use the predicted probabilities from that regression to construct $\hat{Pr}(T = 1|D = d, X)$ for each student.
2. For each half-percentile of the distribution of GPAs (i.e. quantiles 0.5 to 99.5), I implement equation (1.14) to construct the reweighted counterfactual distribution of GPAs in the absence of tracking.
3. I then replicate this process 1000 times on bootstrap resamples of the data, clustering at the dormitory-year level and stratifying by group and period, to construct percentile

bootstrap confidence intervals for the estimated treatment effect at each of the 199 quantiles from step 2.

The Stata code for implementing this estimator is available on request. Note that I do not attempt to estimate the counterfactual minimum and maximum, as inference on these parameters is known to be highly problematic (Horowitz, 2001).

Table 1.13: Estimates of the average treatment effect of tracking within each faculty/school/college using a linear difference-in-differences regression model

| Sample | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|---|--------------------|-------------------|--------------------|-------------------|---------------------|-----------------|-----------------|-----------------------------|
| | All students | | Commerce | Engineering | Law | Medicine | Science | Humanities & Social Science |
| Dependent variable | | | | | | | | |
| Average treatment effect of tracking on the treated | -.123*** (.031) | -.137*** (.03) | -.202*** (.057) | -.196*** (.07) | -1.297*** (.304) | -.022 (.093) | -.028 (.075) | -.128* (.054) |
| Dormitory fixed effects | × | × | × | × | × | × | × | × |
| Individual controls | × | × | × | × | × | × | × | × |
| Faculty fixed effects | | × | | | | | | |
| # dorm-year clusters | 60 | 60 | 60 | 60 | 42 | 43 | 60 | 60 |
| # dormitory students | 6600 | 6600 | 2060 | 1382 | 124 | 678 | 1052 | 1304 |
| # other students | 6685 | 6685 | 1952 | 1004 | 113 | 453 | 824 | 2339 |

Notes: These results show that the negative treatment effect of tracking on student GPAs is present within each of the individual faculties. This suggests that the treatment effect cannot be explained by changes in course-taking behavior. Standard errors in parentheses are clustered at the dormitory-year level, with non-dormitory students are treated as singleton clusters. ***, ** and * denote significance at the 1%, 5% and 10% levels respectively. Observations with missing individual controls are excluded from the sample. All columns control for student gender, language, nationality, race, a cubic in high school graduation test scores and all possible three-way interactions.

Table 1.14: Estimates of the average treatment effect of tracking within the six largest entry-level classes using a linear difference-in-differences regression model

| | (1) | (2) | (3) | (4) | (5) | (6) |
|-------------------------|---------------------------------------|------------------------|--------------------|-------------------|-----------------|-----------------|
| Sample | Economics | Information systems | Management | Physics | Sociology | Statistics |
| Dependent variable | GPA (with zeros for excluded credits) | | | | | |
| ATT of tracking | -.149*** (.045) | -.146** (.071) | -.184*** (.062) | -.263*** (.09) | -.099 (.092) | -.065 (.051) |
| Dormitory fixed effects | × | × | × | × | × | × |
| Individual controls | × | × | × | × | × | × |
| # dorm-year clusters | 59 | 56 | 60 | 59 | 55 | 60 |
| # dormitory students | 2160 | 1174 | 1322 | 822 | 547 | 1801 |
| # other students | 2094 | 1228 | 1296 | 574 | 969 | 1978 |

Notes: These results show that the negative treatment effect of tracking on student GPAs is present within each of the first year entry-level classes. This suggests that the treatment effect cannot be explained by changes in course-taking behavior. Standard errors in parentheses are clustered at the dormitory-year level, with non-dormitory students treated as singleton clusters. ***, ** and * denote significance at the 1%, 5% and 10% levels respectively. Observations with missing individual controls are excluded from the sample. All columns control for student gender, language, nationality, race, a cubic in high school graduation test scores and all possible three-way interactions.

Table 1.15: Instrumental variables estimates of the average treatment effect of tracking using a linear difference-in-differences model, instrumenting the year of college admission with the year of high school graduation and dormitory/non-dormitory status with an indicator for whether the student attended a high school in Cape Town

| | (1) | (2) | (3) | (4) |
|----------------------|-----------------|----------------------|------------------|------------------|
| Sample | All students | Non-missing geocodes | | |
| Dependent variable | GPA | | | |
| Estimator | OLS | OLS | ITT | 2SLS |
| Avg treatment effect | -.129 (.073) | -.114 (.074) | -.104* (.059) | -.144* (.077) |
| # dorm-year clusters | 60 | 60 | 60 | 60 |
| # dormitory students | 7480 | 6187 | 6187 | 6187 |
| # other students | 7188 | 6110 | 6110 | 6110 |

Notes: These results show that the negative treatment effect of tracking on student GPAs is robust to instrumenting for whether each student lives in a dormitory or not and the time period in which they attend the university. Standard errors in parentheses are clustered at the dormitory-year level, with non-dormitory students treated as singleton clusters. ***, ** and * denote significance at the 1%, 5% and 10% levels respectively. Observations with missing high school geocodes are excluded from columns (2) to (4).

Table 1.16: Placebo difference-in-differences tests for differences in the pre-treatment trends for dormitory and non-dormitory students

| | (1) | (2) |
|--|-----------------|----------------------------------|
| Sample | All students | Restricted sample |
| Dependent variable | College GPA | High school grad. test scores |
| Estimator | ITT | |
| Placebo average treatment effect of tracking on the treatment | -.041 (.038) | -.023 (.041) |
| # dormitory students | 6677 | 5457 |
| # other students | 4908 | 4873 |

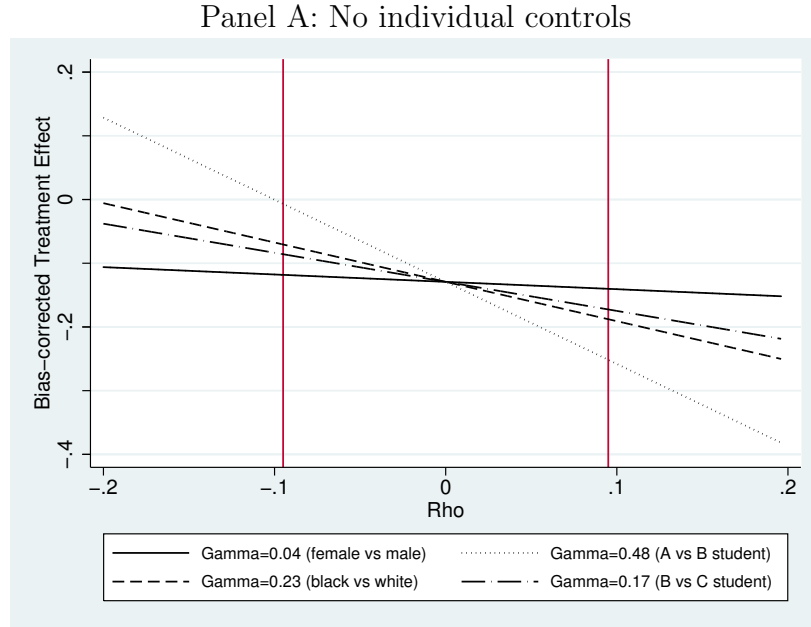
Notes: These results show that the time trends in student GPA and high school graduation test scores are equal for dormitory and non-dormitory students in the pre-treatment period (2001 and 2002 vs 2004 and 2005). Standard errors in parentheses are clustered at the dormitory-year level, with non-dormitory students treated as singleton clusters. ***, ** and * denote significance at the 1%, 5% and 10% levels respectively. Observations with missing individual controls or with missing high school geocodes are excluded from columns (2) to (4).

Table 1.17: Estimated p -values for the test that the average treatment effect of tracking is zero using alternative test procedures in a linear difference-in-differences regression model

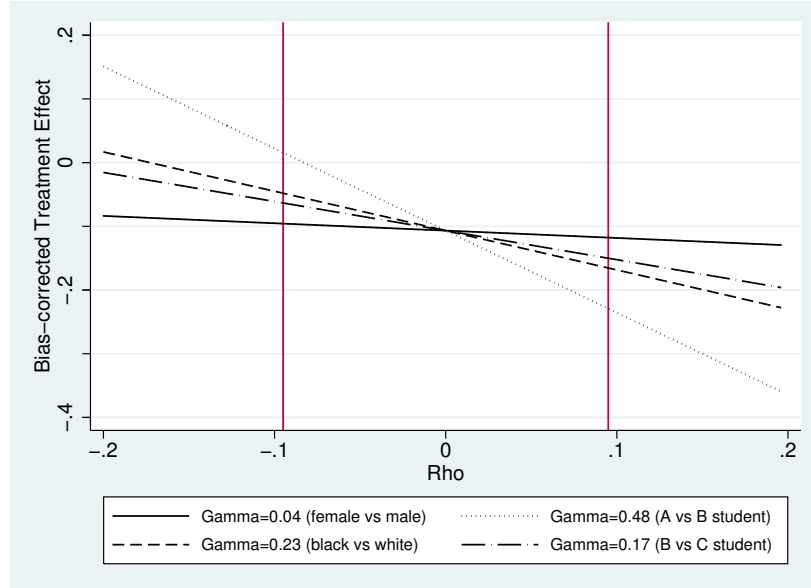
| | (1) | (2) | (3) | (4) |
|---|-------------|-------|-------------------|-------|
| Sample | Full sample | | Restricted sample | |
| Dependent variable | GPA | | | |
| ATT of tracking | -.129 | -.107 | -.113 | -.123 |
| Dormitory fixed effects | | × | | × |
| Year fixed effects | | | × | × |
| Individual controls | | | × | × |
| Panel A: p -values for test of zero treatment effect, dormitory-year clusters | | | | |
| cluster-robust variance estimator | .077 | .078 | .001 | 0 |
| cluster bootstrap variance estimator | .079 | .187 | .001 | .001 |
| cluster bootstrap t -statistic | .014 | .062 | 0 | 0 |
| wild cluster bootstrap t -statistic | .143 | .09 | .005 | .007 |
| Panel B: p -values for test of zero treatment effect, dormitory clusters | | | | |
| cluster-robust variance estimator | .146 | .254 | .001 | .001 |
| cluster bootstrap variance estimator | .155 | .27 | .001 | .001 |
| cluster bootstrap t -statistic | .098 | .191 | 0 | .001 |
| wild cluster bootstrap t -statistic | .266 | .363 | .014 | .012 |

Notes: These results show that the significance of the negative average treatment effect of tracking is robust to alternative test procedures. ***, ** and * denote significance at the 1%, 5% and 10% levels respectively. Observations with missing individual controls are excluded from columns (3) and (4), which control for student gender, language, nationality, race, a cubic in high school graduation test scores and all possible three-way interactions.

Figure 1.8: “Bias-corrected” estimates of the average treatment effect of tracking from a linear difference-in-differences regression model with different assumptions about the relationship between an unobserved scalar and GPA and the time trend in the unobserved variable



Panel B: Controlling for individual and dormitory characteristics



Notes: This figure explores how sensitive the results are to alternative assumptions about potential differences in unobserved student characteristics between the tracking and random assignment periods. The horizontal axis ρ shows “how much” a hypothesized unobserved characteristic differs between the two periods. The vertical lines provide a benchmark by showing the maximum difference between the two periods for any observed characteristic (which is language, as shown in table 1.1.) The γ parameter measures the strength of the relationship between GPA and the hypothesized unobserved characteristic. For example, the solid line indicates the value of γ associated with gender. The figure implies that to entirely explain the observed treatment effect, there would need to be an unobserved student characteristic whose mean changes “as much” from the tracking to the random assignment period as language and which affects GPA almost twice “as strongly” as race.

Chapter 2

How Price Sensitive is School Enrollment? Evidence from Nationwide School Fee Reforms in South Africa

2.1 Introduction

Increasing participation in formal schooling is an important thrust of public policy throughout the developing world. The United Nations' Millenium Development Goals aim to ensure that every child can complete primary education. A wide range of countries have eliminated school fees¹ or given cash grants to households whose school-aged children meet minimum enrollment, attendance or performance thresholds. This policy emphasis is consistent with economic research showing that additional years of schooling increase average individual earnings (Card, 1999; Heckman, Lochner, and Todd, 2006). In particular, some public policies that increase or restrict access to formal schooling in middle income countries have been shown to affect the labor market outcomes of affected individuals (Duflo, 2001; Ozier, 2011). Other studies document that more educated individuals in low income countries adopt new agricultural and health technology faster (Besley and Case, 1993; Dupas, 2013) and have healthier and better educated children (Behrman, Foster, Rosenzweig, and Vashishtha, 1999; Glewwe, 1999). A complementary macroeconomic literature shows a positive association between average years of education and economic growth (Hanushek and Kimko, 2000; Mankiw, Romer, and Weil, 1992). These associations need not all reflect causal relationships but they provide strong suggestive evidence that increased participation in formal schooling improves average economic outcomes.

¹I use the term “school fees” to encompass all mandatory direct payments from households to schools. These may be known as “enrollment fees” or “tuition fees” in different settings. School fees have historically been common in developing country public schools, particularly in the Commonwealth of Nations.

Policymakers have pursued a range of interventions in order to increase participation in formal schooling: “demand-side” interventions include school fee elimination or reduction, conditional cash transfers, and merit scholarships; and “supply-side” interventions include school construction or upgrading, class size reduction, and changes in education technology. The effects of some of these policies on school participation have been thoroughly studied. For example, Fizbein and Schady (2009) review the evidence on conditional cash transfers. They report generally modest effects on school participation that vary substantially with the magnitude and design of the transfers, the characteristics of the target population, and the characteristics of the schooling system. Other interventions have been subject to less detailed study. In particular, very few studies have examined school fee elimination. The effect of this intervention on school participation is a relatively open question.

More than ten African countries have eliminated school fees in all or some primary schools during the past two decades. Several countries are actively considering extending this policy to secondary schools. These interventions may have substantial effects on the education outcomes of school-age youths. Some enrolled youths may be induced to continue their formal education for longer than they otherwise would. Some unenrolled youths may start formal education or return after a period of nonenrollment. The households of inframarginal youths who would enroll in any case will receive an effective lump-sum transfer equal to the school fees they no longer need to pay. Inframarginal youths may also be affected by changes in their class sizes and peer groups induced by rising enrollment. Eliminating school fees changes the financial resources and incentives of schools themselves. If the loss in fee revenue is not offset by compensating transfers from government, schools may need to lay off teachers, cut salaries, or defer investment in physical capital. Even if government transfers make fee elimination revenue-neutral for schools, their accountability shifts in part from students and their families to a potentially more distant layer of government. These effects on school composition, finances, and incentives make fee elimination a potentially very different intervention to conditional cash transfers paid directly to households.

I contribute to the sparse literature on school fees by studying the effect of a geographically targeted school fee elimination intervention in South Africa. The national government eliminated school fees for schools located in high-poverty neighborhoods between 2006 and

2007. Specifically, schools were ranked based on the poverty rate in the surrounding neighborhood and then divided into five quintiles. Schools in the first and second quintiles were required to eliminate fees and were given additional per-student government transfers intended to offset the lost fee revenue. This generated time-series variation in school fees – before and after the intervention – and controlled cross-sectional variation – high- versus low-poverty quintiles. I use this variation to generate difference-in-differences and regression discontinuity estimates of the effect of school fee elimination on enrollment and other education outcomes. This approach is similar to prior studies of fee elimination in Colombia (Barrera-Osorio, Linden, and Urquiola, 2007) and South Africa (Borkum, 2011). The estimates are less subject to concerns about confounding than those based on time-series variation alone (Deininger, 2003) or cross-sectional and time-series variation induced by the incidence of civil war (Fafchamps and Minten, 2007). My primary analysis uses a restricted access school-by-grade-by-year longitudinal dataset collected by South Africa’s national Department of Education.²

I show that fee elimination has a relatively small effect on enrollment. My preferred estimates suggest that 1 to 3 additional students are induced to enroll in the average school. This increases the baseline enrollment level by 1% or less and the baseline enrollment rate by less than 1 percentage point. The data do not allow me to identify the price elasticity with respect to baseline fee levels but informal calculations imply it is smaller than -0.01 (in absolute value). The effects are robust to accounting for pre-treatment time trends and pre-treatment differences between high- and low-poverty schools’ characteristics. They are not driven by net transfers of students between treated schools and those that continued to charge school fees, even though South Africa allows substantial school choice. The increased enrollment occurs entirely in secondary grades, partly due to the very high baseline enrollment rate for primary school-aged youth. The zero effects on primary grades suggest that schools are not systematically overreporting enrollment in order to obtain more per-student government transfers.

These results imply that demand for schooling is relatively price insensitive in the neigh-

²The dataset includes all schools in the country but key variables are missing for schools in four of the nine provinces. My analysis is therefore restricted to the Eastern Cape, Gauteng, KwaZulu-Natal, Northern Cape and Western Cape provinces.

borhoods treated by the fee elimination intervention. I explore several possible explanations for this pattern. The enrollment response remains small in populations with low baseline enrollment, ruling out ceiling effects on enrollment as a primary explanation. Schools do not appear to face binding capacity constraints on enrollment. Enrollment effects are larger in populations less likely to face credit constraints, so pecuniary costs other than school fees may still inhibit enrollment. However, post-treatment enrollment rates in populations less likely to be credit constrained remain well below one. These results suggest that price insensitive demand in part reflects low valuation of additional years of enrollment amongst unenrolled youths and their parents. Low valuation may reflect high opportunity costs or low returns to education. Labor market opportunity costs are likely to be lower in South Africa than in other developing countries: youth employment is exceptionally high (Banerjee, Galiani, Levinsohn, McLaren, and Woolard, 2008) and the smallholder agriculture sector is unusually small (Terreblanche, 2002). This leaves non-pecuniary opportunity costs or low returns to education as explanations. Education “quality,” measured by standardized test scores, grade progression rates, and high school graduation is low for already-enrolled students at schools located in high poverty neighborhoods (Lam, Ardington, and Leibbrandt, 2010; van den Berg and Louw, 2007). Unenrolled youths may perceive that enrollment will not lead to grade progression or learning, which would explain the small enrollment effects of fee elimination.

Eliminating school fees may also affect school finances and resources, which may affect enrollment behavior by forward-looking youths. I show that some simple measures of school resources and performance – class size, grade progression and dropout rates, and the socioeconomic status of the student body – are largely unaffected by fee elimination. Students may have been deterred from enrolling by concerns about falling school resources, but such beliefs would not be consistent with the outcomes of the intervention.

In sum, these results suggest that nationwide geographically targeted school fee elimination had small effects on student enrollment and on measured school characteristics. The intervention’s largest effect may have been on households containing inframarginal students, who no longer needed to pay fees. The net welfare effect of the program depends on the weight assigned to these generally low-income households relative to the taxpayers who

funded the intervention.³ The net welfare effect is more likely to be positive if some students with a high valuation enrollment were previously deterred from enrollment by credit constraints. Edmonds (2006) supports this consideration by showing that some rural households in South Africa were credit constrained with respect to school enrollment. Finally, the absence of substantial student transfers from fee-charging into fee-eliminating schools shows that geographic targeting of pro-poor subsidies can be accurate under certain circumstances. This may be a cost effective alternative to income-dependent pricing of public services when verifying individual income is difficult. However, this result may not generalize to situations where demand for public services is more price sensitive.

Organization of the paper: I develop a simple conceptual framework in section 2.2 that motivates the empirical analysis. Section 2.3 provides an overview of the South African education system and fee elimination intervention. I then explain how the design of the intervention motivates the identification strategies. Section 2.4 reports the treatment effects of school fee elimination on school-level enrollment. I then discuss the magnitude of these effects and a number of robustness checks. I explore a number of explanations for these price-insensitive demand estimates in section 2.5. Section 2.6 reports treatment effects on measures of school composition, performance and resources. I also consider how these results relate to the enrollment effects. Section 2.7 concludes.

Related literature: This paper directly advances a growing literature that studies the effect of school fee eliminations on student enrollment in primary and secondary schools. Studies in Madagascar (Fafchamps and Minten, 2007), Malawi (Al-Samarrai and Zaman, 2000), Kenya (Lucas and Mbiti, 2009), and Uganda (Deininger, 2003) have found large increases in enrollment of up to 100% off relatively low bases. These results are typically an order of magnitude larger than those I find, which may reflect more elastic education demand in lower income countries or limitations of these research designs. Most studies rely on simple comparisons of enrollment before and after a nationwide fee elimination. The implementation of other simultaneous policy changes may also have affected enrollment and resulted in upward biased estimates of the effect of fee elimination.

³The intervention was funded by the national Department of Education out of the general fiscus. The incidence of the cost of the intervention cannot be directly determined. Most national government revenue in South Africa is raised through a progressive personal income tax and flat rate corporate income tax.

Barrera-Osorio, Linden, and Urquiola (2007) and Borkum (2011) provide perhaps the most credible evidence to date on this question. The former paper examines a natural experiment in Bogota, Colombia, where the local government reduced school fees for households whose socio-economic status fell below a threshold level. The authors find that enrollment increased by 0-5 percentage points in households just below the threshold, relative to households just above the threshold who did not qualify for fee reductions. The latter paper also studies the South African fee reform in a single province using more limited data and finds a rise in enrollment of 0-2%.

The contrast between these two strands of the literature may reflect a number of differences. First, the interrupted time series and panel data used in the former literature may be subject to substantial upward biases due to correlated policy changes. Second, the regression discontinuity designs used in the latter literature and in my own work estimate valid treatment effects only in the neighborhood of the cutoff. If poorer households are more responsive to fee eliminations, treatment effects may be considerably larger well below the cutoff. I find some evidence in my data that increases in enrollment after school fee elimination are negatively correlated with neighborhood poverty rates. Third, countries studied by the latter literature typically have much higher baseline enrollment rates so the treatment effects may be restricted by ceiling effects.

This paper is also related to a substantial literature on the enrollment effects of conditional cash transfers. These transfers have typically been offered in countries without school fees and are designed to offset the lost contributions to home production and labor market earnings from enrolled children. These programs typically raise enrollment by less than 10%, though there is some heterogeneity across studies (Angelucci and di Giorgi, 2009; Glewwe and Kassouf, 2011; Schultz, 2004; Todd and Wolpin, 2006). This heterogeneity has been ascribed to differences in baseline enrollment rates across different countries and to differences in the extent to which conditions bind. For example, some cash transfers are only conditional on school enrollment, while others also impose conditions on attendance or grades. All else equal, the latter tend to have smaller effects, perhaps because the stricter conditions render more potential students inframarginal with respect to the intervention (Filmer and Schady, 2008; Kremer, Miguel, and Thornton, 2009). The effect sizes that I find from eliminating

school fees are roughly comparable to those associated with conditional cash transfers in Brazil and Mexico, suggesting that the two policies may have relatively similar effects when applied in countries with comparable levels of baseline enrollment.

A small number of studies have considered the effect on school participation of reducing other pecuniary costs of education. Evans, Kremer, and Ngatia (2009) show that distributing free school uniforms increased attendance amongst already-enrolled youths. In contrast, Hidalgo, Onofa, Oosterbeek, and Ponce (2013) find negative effects of free school uniforms on attendance, though this may have been driven by very low treatment compliance.

There is also a small literature that studies the enrollment effects of changes in the cost of education in the developed world. See Dynarski (2003), Dynarski, Gruber, and Li (2009), Kane (1994), and Seftor and Turner (2002) for examples and Neal (2002) for a discussion of the challenges faced by this research agenda. However, these research designs typically focus on margins different to those in the development literature: the choice between private and (free) public primary or secondary education and the choice between different types of postsecondary institutions with different costs. These are very different to the choice between fee-charging public and no primary or secondary education, which is the more relevant margin in much of the developing world. While these studies provide useful insights into the nature of education investment decisions, they do not reduce the need for better evidence on the relationship between school fees and enrollment.

2.2 Theoretical Framework

Consider a simple reduced form model of the enrollment decision. Assume that each youth i decides whether to enroll in the local school s at time t . Panel A of figure 2.1 depicts this decision in a simple demand and supply framework. The vertical axis V shows the value and cost of enrolling for an additional year of schooling in money metric terms. The horizontal axis shows the proportion $P \in [0, 1]$ of youth in this school's neighborhood who are enrolled. The downward-sloping demand curve D captures the idea that the value of enrollment is heterogeneous across individuals within a school.⁴ This heterogeneity may arise

⁴Throughout this section I assume that the demand curve has no discontinuities in the interior. This follows from the assumption that the density of individual valuations of enrollment within a school is strictly

from different levels of academic ability, different outside options, and the fact that students have reached different grade levels at time t . Assume for now that the cost of enrollment is identical for all agents, so that the cost curve C is horizontal. This assumption is relaxed below. Equilibrium enrollment occurs at p_1 . Note that v_1 is the net value of enrollment for the marginal student (gross value less cost), not a market-clearing price.

Eliminating school fees shifts the cost curve downward from C_1 to C_2 . The new equilibrium enrollment rate is $p_2 \geq p_1$. Students can be divided into three categories with respect to the fee elimination intervention. p_1 of the students are inframarginal and enroll whether fees are charged or not; $1 - p_2$ of the students are inframarginal and do not enroll even if fees are not charged; $p_2 - p_1$ of the students are marginal and enroll if and only if fees are eliminated. This yields two general implications of the framework:

- I1** Eliminating school fees increases total enrollment and the enrollment rate. Larger baseline school fees will lead to a larger effect on enrollment. If the baseline enrollment rate is at or near one, eliminating fees will have a smaller effect on enrollment.
- I2** The change in the enrollment rate equals the proportion of students whose gross value of enrollment is smaller than the cost of enrollment including fees C_1 but larger than the cost of enrollment excluding fees C_2 .

This simple framework treats enrollment as a static decision, rather than adopting a dynamic discrete choice framework. The framework also abstracts away from uncertainty. In practice, grade progression is far from universal in South Africa and the potentially stochastic relationship between enrollment and grade progression may vary across schools and individuals within a school (Lam, Ardington, and Leibbrandt, 2010). The value of enrollment V can be interpreted as the expected present value of the discounted stream of future benefits from enrollment, which implicitly includes the option value of enrolling in higher grades in future years. Explicitly accounting for dynamics and uncertainty does not appear to generate any empirical predictions that can be implemented with my data.

Panel B of figure 2.1 considers two potential valuation curves: D^X is relatively elastic in the neighborhood of the equilibrium and inelastic D^Y is relatively inelastic. Eliminating

continuous.

fees increases the equilibrium enrollment rates from p_1^X and p_1^Y to p_2^X and p_2^Y . The former change is clearly larger, generating another implication of the framework:

- I3** The effect of fee elimination on enrollment is increasing in the local elasticity of the valuation curve D .

Panel C of figure 2.1 considers three potential demand curves: convex D^X , linear D^Y , and concave D^Z . The change in enrollment rates induced by eliminating fees is largest for D^Y and smallest for D^Z . The mean valuation amongst the “always-enrollers” is largest for D^Z and smallest for D^X . This generates another implication of the framework:

- I4** There is not a monotonic relationship between the mean valuation amongst students enrolled when school fees are charged and the increase in enrollment rate induced by eliminating fees.

I4 is particularly empirically important. It implies that school-level treatment effects of school fee elimination will not necessarily be correlated with proxies for the valuation of enrollment prior to elimination. Hence, enrollment may rise by a smaller or larger margin in “good” than “bad” schools. In this depiction, the proportion of always-enrollers and their mean valuation of enrollment are positively correlated. However, this does not hold more generally.

The framework can be extended to allow for a number of more realistic elements. In particular:

- I5** If baseline costs of enrollment are heterogeneous, the cost curve will be upward-sloping and the effect of fee elimination on enrollment will be attenuated relative to the constant-cost case.
- I6** If schools face binding capacity constraints, the equilibrium enrollment rate will be constrained to be below 1 and the effect of fee elimination on enrollment may be attenuated.
- I7** If fee elimination reduces the valuation of enrollment, perhaps due to resource constraints or negative peer effects, the demand curve will shift downward and attenuate or potentially reverse the effect of fee elimination on enrollment.

This framework generates several guidelines for the empirical analysis. First, eliminating fees should increase the enrollment rate (I1). Second, if the enrollment rate falls, this must be due to negative effects of fee elimination on the perceived valuation of enrollment (I7). Third, if the change in the enrollment rate is small, this may reflect near universal baseline enrollment (I1), a low initial value of fees (I1), inelastic demand (I2/3), heterogeneous costs (I5), binding capacity constraints (I6), and/or negative effects on the valuation of enrollment (I7). Fourth, the magnitude of the treatment effects need not be explained by the mean valuation of enrollment amongst students enrolled at baseline (I4).

The framework as written assumes that credit constraints never bind on the enrollment decision. The presence of credit constraints will have an ambiguous impact on the relationship between fee elimination and enrollment. In terms of figure 2.1 panel A, eliminating fees will induce some credit-constrained students with valuations above v_1 to enroll. However, credit-constrained students with valuations between v_2 and v_1 will not be able to enroll. If the former group is larger than the latter, the treatment effect of fee elimination on enrollment will be increased relative to a world with no credit constraints.

2.3 Background Information and Identification Strategy

2.3.1 Background on South African Education

South Africa is a middle income country with a history of sharp economic, political, and social inequality. The public education system was racially segregated until the early 1990s, and per capita government expenditure on white schools was orders of magnitude larger than on black schools. Both enrollment and high school graduation rates differed sharply by race and there is some evidence of large differences by household socio-economic status (Fedderke, Luiz, and de Kadt, 2000; Seekings and Nattrass, 2005). A small number of black students from low income households enrolled in private schools, mostly church-run. Despite recent growth in the number of small, for-profit private schools in low income communities, the best available data suggest that the private sector remains negligible relative to the public sector and considerably smaller than in South Asian countries (Centre for Development Enterprise, 2010).

State expenditure on black education rose substantially in the 1970s, 1980s, and 1990s and this was associated with rapidly rising enrollment rates. However, substantial evidence suggests that the quality of education remained very low in historically black schools. Curricula at these schools had deliberately focused on non-academic subjects until the 1990s, reflecting the *apartheid* government's insistence on preparing black students for manual employment only. Few students completed secondary schooling, pass rates on high school graduation examinations were low, and even fewer students took mathematics or physical science as high school subjects (Fedderke, Luiz, and de Kadt, 2000). The education system was officially desegregated in the early 1990s but the majority of black students continue to attend historically black schools. Sharp racial differences are visible in which schools students attend and whether they attend school at all, with black enrollment up to 10 percentage points lower than white enrollment.

There are two approaches to estimating baseline enrollment rates and they produce slightly different results. The first approach divides the total number of students that schools report enrolling by the population projected from census data. This method yields national enrollment rates of 85% and 78% for 7-13 and 14-18 year-old youths respectively. The second approach uses nationally representative household survey data and yields enrollment rates of 96% and 85% in these age brackets.⁵ Both approaches suggest that enrollment is high at younger ages and tapers into adolescence. These are *net enrollment rates*, which measure the proportion of the age-eligible population enrolled. This may overstate the enrollment rate in a country with substantial rates of late school entry, grade repetition, and temporary dropout (Department of Education, 2009).

These baseline enrollment rates show that there was substantial room for fee elimination to increase enrollment. I am not aware of any nationally representative dataset that would allow calculation of the baseline enrollment rate in neighborhoods where fees were eliminated. There is a strong negative correlation between youth enrollment and household socio-economic status, suggesting that baseline enrollment was substantially lower in

⁵These data are reported in Department of Education (2009) and Department of Education (2011). The survey data are drawn from the General Household Survey, a nationally representative annual project conducted by Statistics South Africa. The survey is relatively comparable to its namesake in the United Kingdom.

treated than in untreated neighborhoods. The difference between enrollment behavior in low- and high-income households led some policymakers to advocate interventions to reduce the pecuniary cost of schooling (Pampallis, 2008). Such policies aligned closely with the post-*apartheid* government’s long-standing stated commitment to free education. They may also have been motivated by widespread primary school fee eliminations in other African countries during the 1990s and 2000s. Edmonds (2006) shows that rural households that received a fully anticipated income increase (a state old age pension) were more likely to enroll their children than those that did not. This is consistent with credit constrained enrollment decisions.

Low income and credit constraints are not the only reasons offered for low enrollment rates in some neighborhoods. Dropout in secondary school has also been ascribed to the low quality of schools in low income neighborhoods and may be a rational response to low returns to education in these schools. There are no direct measures of returns to education at different types of schools, so these arguments typically infer low returns to education from evidence of low school quality. While measuring “school quality” is a difficult process, it is true that South African students attending schools in low-income neighborhoods perform considerably worse on international literacy and numeracy assessments than poorer students from other African countries (van den Berg and Louw, 2007). South Africa has a system of nominal school choice, under which individuals from low income neighborhoods can in principle enroll in schools in high income neighborhoods. The limited data available on this phenomenon suggests that it is uncommon. This may reflect a combination of high commuting costs in cities that are still highly segregated by income and race, and social and cultural barriers that limit low income students’ ability to integrate into schools in high income neighborhoods.

Two interventions were introduced in order to reduce the pecuniary cost of education and promote higher enrollment: means-tested *individual-level school fee waivers* and *school-level school fee eliminations*. The former intervention was introduced in 1996 and required that schools grant partial fee waivers to any household that either earned less than 10 times the per-student school fee or was eligible to receive a means-tested government child grant. The latter requirement meant that a large proportion of the country’s students were eligible for the

waiver. However, household survey data from 2005 show that approximately 2% of students benefitted from fee waivers and some media reports suggest that parents were discouraged by schools from requesting waivers (Borkum, 2011; Hanes, 2006). This intervention reduced the actual price of education below its nominal level and so risks attenuating any effect of the second intervention on enrollment behavior. I argue below that any bias is likely to be small, due to the rarity of fee waivers and the discontinuous implementation of the fee elimination intervention.

2.3.2 *School Fee Elimination*

The school fee elimination intervention was announced in 2006 and implemented in 2007.⁶ Schools treated by this intervention were required to eliminate all tuition and enrollment fees, though the status of additional fees for extra-curricular activities was not regulated. These “no fee” schools were chosen by a complex three-stage interaction between provincial and national governments, laid out in guidelines published by the national Department of Education.

In the first stage, provincial governments assigned each school in their province a “poverty score” based on characteristics of the electoral ward in which it was located.⁷ These scores ranked all schools within the province from least to most poor, with ties permitted. The national Department of Education provided each province with ward-level data on income, employment, education, health, and “living environment” from the 2001 census as a starting point for the assignment of poverty scores. Provinces were permitted choose their own weighting of these five data series and to make *ad hoc* adjustments to the resultant score based on within-ward heterogeneity. They were not permitted to use any data collected directly from schools, such as administrative data on school’s physical facilities or student-teacher ratios. Wildeman (2008) conducted anonymous interviews with provincial officials responsible for creating the poverty scores and reported that most of the *ad hoc* adjustments were made for schools near the boundaries of electoral wards, as the socio-economic characteristics of their students may have differed from those of the electoral ward. Wilde-

⁶South Africa’s academic year runs from January to December.

⁷The electoral ward is not an administrative unit in South Africa. Assignment took place at this level because it is the smallest geographic unit at which census data is available.

man’s interviewees reported no incidents of schools lobbying provincial officials to change their scores, although lobbying may have occurred in the third stage described below. The formulae used to determine the poverty scores were left to the discretion of the provinces and no province has made its formula publicly available.⁸

In the second stage, the national government divided all schools in the country into five “quintiles” based on these poverty scores. Each quintile was intended to contain approximately 20% of the students in the country (based on 2006 enrollment data) and each quintile would contain approximately “equally poor” school neighborhoods in each province. So in the relatively poor Eastern Cape province, 35% and 6% of all schools were assigned to the first and fifth quintiles respectively; in the relatively low poverty Western Cape province, 7% and 23% of all schools were assigned to the first and fifth quintiles respectively. The choice of how many schools were to be treated in each province was based on province-level data from the 2001 census but the exact algorithm used for this decision is unclear. All schools in quintiles 1 and 2 were intended to be no fee schools. By determining the number of schools to be treated in each province, the national government implicitly specified a cutoff value of the poverty score above which all schools were the “intention to treat” and below which all schools were the “intention to control” group. National government officials report that the number of schools to be treated was chosen after the poverty scores had already been assigned, so it was not possible for poverty scores to be precisely manipulated in the neighborhood of the cutoff.

In the third and final stage, provincial governments decided which schools were to abolish fees, which created an “actual treatment” group of schools. The actual treatment assignments followed the intended treatment assignments relatively closely: 98% of quintile 1/2 schools above the cutoff eliminate fees while only 6% of quintile 3/4/5 schools below the cutoff do so. The discrepancies may reflect lobbying by schools above the cutoffs who wished to continue charging fees or by schools below the cutoff who wished to eliminate them. The frequency of these discrepancies varies across provinces: 30% of schools have

⁸Each province used a different scale for the poverty scores. I standardize these by recentering them at the cutoff between quintiles 2 and 3 and rescaling them to have standard deviation one within each province. The results are reasonably robust to alternative standardizations: rescaling the range to one within each province or rescaling the variance to minimize the sum of the differences between the quintile 1/2 cutoff and the quintile 3/4 cutoff.

different intended and actual treatment statuses in the least compliant province (Northern Cape), while intended and actual treatment statuses are identical for all schools in one other province (Gauteng).

2.3.3 Identification Strategy

The design of the fee elimination intervention makes it a natural candidate for analysis using both difference-in-differences and regression discontinuity methods. I compare schools below the cutoffs, the intended control group, with schools above the cutoffs, the intended treatment group. If the poverty scores are “as good as randomly assigned” (Lee and Lemieux, 2010) in the neighbourhood of the cutoffs, these two groups differ only in their treatment status and so any differences in enrollment between the two groups may be interpreted as a causal effect of the fee elimination intervention.

Baseline difference-in-differences specification: My primary estimation sample consists of schools in quintiles 2 and 3. I use quintile 2 schools as the “intention to treat” group and quintile 3 schools as the control group. By eliminating quintile 1, 4, and 5 schools, I restrict the sample to schools with relatively similar baseline characteristics in line with the spirit of the regression discontinuity design. I begin by estimating

$$\text{Fee elimination}_i = \alpha_0 + \alpha_1 \text{High poverty score}_i + \epsilon_i \quad (2.1)$$

by ordinary least squares. The coefficient α_1 captures the difference in the probability of fee elimination between schools with high poverty scores (i.e. in quintile 2) and schools with low poverty scores (i.e. in quintile 3). This tests whether assignment to treatment is broadly consistent with the process described in the previous subsection. I use the cross-section of schools in 2007 to estimate this model.

I then estimate

$$\begin{aligned} \text{Enrollment}_{it} = & \beta_0 + \beta_1 \text{High poverty score}_{it} + \beta_2 \mathbf{1}\{\text{Year} \geq 2007\} \\ & + \beta_3 \text{High poverty score}_{it} \times \mathbf{1}\{\text{Year} \geq 2007\} + \nu_{it} \end{aligned} \quad (2.2)$$

The coefficient β_3 captures the difference in the change in enrollment from pre-2007 to post-

2007 between schools with high poverty scores (i.e. in quintile 2) and schools with low poverty scores (i.e. in quintile 3). This tests whether schools in the intention to treat group experience a larger increase in enrollment from 2005/6 to 2007/8 than other schools. I use a panel of school-level enrollment between 2005 and 2008 to estimate this model and I use a cluster-robust variance estimator that allows unrestricted intertemporal correlation in ν_{it} for each school i .⁹

I finally estimate

$$\begin{aligned} \text{Enrollment}_{it} = & \gamma_0 + \gamma_1 \text{Fee elimination}_{it} + \gamma_2 \mathbf{1}\{\text{Year} \geq 2007\} \\ & + \gamma_3 \text{Fee elimination}_{it} \times \mathbf{1}\{\text{Year} \geq 2007\} + \eta_{it} \end{aligned} \quad (2.3)$$

using instrumental variables, with the indicator for fee elimination instrumented by high poverty score indicator and the interaction term treated analogously. The coefficient γ_3 captures the difference in the change in enrollment from pre-2007 to post-2007 between no fee schools and fee charging schools who comply with their intended school fee policy. This captures the effect of fee elimination on enrollment for schools that comply with their intended treatment status. I again use a panel of school-level enrollment between 2005 and 2008, with a cluster-robust variance estimator. The results in section 2.4 verify that the instruments used in estimating equation (2.3) easily pass the appropriate tests for instrument strength.

Identification of $(\alpha_1, \beta_3, \gamma_3)$ relies on the assumption that the counterfactual trend in enrollment for quintile 2 and quintile 3 schools from 2005/6 and 2007/8 would have been identical if fees had not been eliminated. This assumption may be problematic if the two groups of schools differ on observed or unobserved characteristics that are associated with enrollment trends. If such “confounding” occurs, the identification assumption will fail. I use two strategies to address this potential concern.

Rewighted difference-in-differences specification: I first consider the possibility that quintile 2 (intention to treat) and quintile 3 (control) schools may differ on observed

⁹I pool 2005 and 2006 together and 2007 and 2008 together to smooth out potential measurement error in the reported enrollment data. Formal tests do not reject equality of enrollment in 2005 and 2006 and in 2007 and 2008.

characteristics. The panel structure of the difference-in-differences design is equivalent to including school-level fixed effects in equations (2.1) – (2.3). This accounts for any differences in the level of enrollment between intention to treat and control schools. However, it does not account for the possibility that enrollment trends may be systematically correlated with baseline observed characteristics that differ systematically between intention to treat and control schools. I observe data on a vector of baseline characteristics from 2005 and 2006: enrollment, number of grades offered, phase (primary, intermediate, or secondary), location (urban or rural), historical racial classification (Asian, black, white, mixed race, or founded after desegregation), designation as a mathematics and science specialization school, partial self-governance status, class size, student-teacher ratio, proportions of part-time and temporary teachers, dropout rate, grade promotion rate, proportion of orphans in the school, and proportion of students in the school whose families receive government social grants. The means of almost all of these characteristics differ significantly between quintile 2 and 3 schools and the χ^2 test statistic for equality of all characteristics is 2323. I therefore construct a sample of control schools weighted to have the same distribution of observed characteristics as the intention to treat schools. I follow Abadie (2005) and DiNardo, Fortin, and Lemieux (1996) in using the reweighting function

$$\omega(X_{it}) = \frac{Pr(\text{High poverty score}_i = 1|X_{it})}{1 - Pr(\text{High poverty score}_i = 1|X_{it})} \quad (2.4)$$

for all schools in the control group. This term assigns high weight to control schools whose observed characteristics X_{it} in the baseline period (2005 and 2006) “look like” those of the intention to treat schools.¹⁰ I estimate the predicted probability using a logistic regression of an indicator for high poverty scores (i.e. quintile 2 schools) on the full vector of observed characteristics and quadratic terms in the continuous variables.¹¹ I then estimate equations

¹⁰I include missing data indicators where values of any of the observed characteristics are not reported in the dataset.

¹¹Although the two groups differ on average characteristics, there is reasonable overlap in the predicted probabilities $\hat{Pr}(\text{High poverty score}_i = 1|\text{High poverty score}_i = 1, X_{it})$ and $\hat{Pr}(\text{High poverty score}_i = 1|\text{High poverty score}_i = 0, X_{it})$. The maximum predicted probabilities in the two samples are almost identical and less than 1% of the control observations have predicted probabilities below the minimum in the intention to treat group.

(2.1) – (2.3) using weighted least squares.¹²

After constructing the weights, I test whether the means of the observed characteristics differ between the intention to treat and the reweighted control group. I fail to reject the null hypothesis of equal means for any of the individual observed characteristics and fail to reject the null of joint equality (χ^2 test statistic of 28.3 with p -value 0.45) from a weighted seemingly unrelated (SUR) regression. This confirms that the reweighted sample of control schools are statistically indistinguishable from the intention to treat schools on observed baseline characteristics. Hence, the weighted least squares estimates from equations (2.1) – (2.3) will be purged of any confounding due to differences in enrollment trends correlated with observed school-level characteristics.

Regression discontinuity differences specification: The strategy above will account for any confounding due to differences in observed baseline characteristics between intention to treat and control schools. However, it will not account for confounding due to differences in *unobserved* characteristics that are correlated with enrollment trends between the two groups. It will also be problematic if enrollment trends are correlated with socio-economic status as measured by poverty scores. These determine schools' status in the intervention and so are disjoint between intention to treat and control schools and cannot be used in the reweighting algorithm. I address this concern by taking advantage of the process used to assign schools to treatment status. Specifically, I estimate

$$\begin{aligned} \text{No fee school}_i = & \mathbf{1}\{\text{Poverty score} \geq 0\} \times f^+(\text{Poverty score}_i) \\ & - \mathbf{1}\{\text{Poverty score} < 0\} \times f^-(\text{Poverty score}_i) + \epsilon_i \end{aligned} \quad (2.5)$$

and

$$\begin{aligned} \Delta \text{Enrollment}_i = & \mathbf{1}\{\text{Poverty score} \geq 0\} \times f^+(\text{Poverty score}_i) \\ & - \mathbf{1}\{\text{Poverty score} < 0\} \times f^-(\text{Poverty score}_i) + \epsilon_i \end{aligned} \quad (2.6)$$

where f^+ , f^- , g^+ , and g^- are polynomial or local linear functions of poverty scores. The idea

¹²I approximate the standard errors of the resultant estimators using 100 replications of a bootstrap algorithm that iterates over both stages of the estimation: construction of the weights and estimation of the linear difference-in-differences models. The bootstrap algorithm resamples school clusters with replacement.

behind these models is to specify flexibly the relationship between the outcome of interest (respectively, no fee status and the change in enrollment from 2005/6 to 2007/8). I can then evaluate

$$\lim_{\text{Poverty score} \downarrow 0} \hat{f}^+(\text{Poverty score}_i) - \lim_{\text{Poverty score} \uparrow 0} \hat{f}^-(\text{Poverty score}_i)$$

to estimate the magnitude of the change in the outcome of interest that occurs in the neighborhood of the cutoff poverty score that separates intention to treat and control schools. Provided schools are not able to manipulate precisely the poverty score they are assigned by province, untreated schools on one side of the cutoff should be a valid counterfactual for treated schools on the other side of the cutoff. Note that I use a time-differenced outcome variable in equation (2.6) so the specification already removes any time-invariant observed or unobserved characteristics. This identification strategy may be considered “doubly robust” relative to standard regression discontinuity designs that use only cross-section data.

This identification strategy generates two testable predictions. First, I verify that none of the observed characteristics listed above have statistically significant or economically meaningful “jumps” at the cutoff. Second, I verify that there is no evidence that the density of the poverty score variable jumps at the cutoff (see figure 2.2). McCrary (2008) notes that such a jump would be consistent with manipulation of the poverty scores in order to control treatment assignment.

2.4 Effect of fee elimination on enrollment

Figure 2.3 shows the time trend in enrollment rates calculated from the General Household Survey. Comparing 2006 and earlier years to 2007 and subsequent years suggests that enrollment amongst primary school-aged youths was largely unaffected by the intervention, while enrollment amongst secondary school-aged youths rose very slightly. Table 2.1 presents the more formal difference-in-differences analysis comparing quintile 2 (intention to treat schools) to quintile 3 (control schools). Column 1 shows that quintile 2 schools are 94 percentage points more likely to eliminate school fees than quintile 3 schools. This high rate of compliance with the intervention means that poverty score is a strong instrument (first stage F -statistic over 50000) for fee elimination. Column 2 shows that enrollment in schools that

eliminated fees rose from 2005/6 to 2007/8 by 3.6 students more on average than in other schools. The 95% confidence interval for this intention to treat estimate is 0 to 7.1 students and the point estimate is marginally significantly different to zero. The corresponding instrumental variables estimate in column 3 is 3.8 students per school (95% confidence interval 0 to 7.5).

This effect size can be expressed in several different metrics. First, it directly measures the number of additional students induced to enroll by the intervention in each school. Second, the intention to treat and instrumental variables estimates respectively imply 0.9% and 1% increases in total baseline enrollment (confidence intervals respectively 0 to 1.8% and 0 to 1.9%). Third, the effect can be converted into a change in the enrollment rate under some additional assumptions. The baseline enrollment rate for youth aged 7-18 in 2005/6 in the General Household Survey was 92%. If this enrollment rate applied to the neighborhoods around fee-eliminating schools, the intervention would increase the enrollment rate by 0.83 (ITT) to 0.91 (IV) percentage points.¹³ The actual enrollment rate is strongly negatively correlated with household socio-economic status and is likely to be considerably lower near treated schools. Hence, the change in the enrollment rate for any given change in the level of enrollment will be smaller. The upper bound of the 95% confidence intervals allow me to rule out effect sizes larger than 1.7 percentage points for the intention to treat estimate or 1.8 percentage points for the instrumental variables estimates. This implies that a minimum of 6% of youths remained unenrolled despite the elimination of fees, with the rate in affected neighborhoods probably substantially higher.

Converting this effect size into an elasticity is complicated by the lack of available data on baseline school fees. The South African Department of Education has collected these data for all schools around the country but I have not yet been able to obtain access to them. Without these data, it is not possible to assign a monetary value to the eliminated fees or to calculate how this compares to other pecuniary costs of enrollment (transport, uniforms, books, etc.). I use two sources of survey data to obtain a rough value of the elasticity. First, the proportion of youths aged 7-18 in the General Household Survey who were enrolled and

¹³This follows from multiplying the baseline enrollment rate by the percentage increase in the baseline enrollment level.

paying school fees fell from 98% in 2005/6 to 73% in 2007/8. Although these respondents cannot be matched to schools and hence to treatment status, I assume that the proportion of students in each quintile is equal and that all students not paying fees at baseline were in the treated quintiles 1 and 2. This implies a baseline fee payment rate of 95%. If all students who paid fees after the intervention were in the control quintiles 3 - 5, then the fee payment rate in treated schools fell to 32%.¹⁴ This implies that the arc price elasticity of enrollment with respect to paying any fees is

$$\begin{aligned}\epsilon_1 &= \frac{\Delta \text{Enrollment rate}}{\Delta \text{Fee payment rate}} \times \frac{\text{Fee payment rate}}{\text{Enrollment rate}} \\ &= -\frac{0.0083}{0.63} \times \frac{0.6 \times 0.95 + 0.4 \times 0.32}{0.92} \\ &= -0.010\end{aligned}$$

for the intention to treat estimate and -0.011 for the instrumental variables estimate. Even using the upper bound of the 95% confidence intervals for the change in the enrollment rate yields elasticities of -0.019 and -0.020.

The second approach to calculating an elasticity uses calculations by Branson, Lam, and Zuze (2012) based on data from the National Income Dynamics Survey.¹⁵ They report that the average school fees paid by enrolled youths of any age in 2007 were 64 rands¹⁶ in quintile 1/2 schools and R301 rands in quintile 3/4 schools. Under the extreme assumption that these values were identical prior to the fee elimination, the intervention reduced fees by 79% of their baseline value. If fees were the only pecuniary cost of education, this implies that

¹⁴This calculation abstracts away from the fact that the number of enrolled students is changed by the intervention. The small magnitude of this change means that the results are robust to taking this into account.

¹⁵The National Income Dynamics Survey is a nationally representative panel dataset that began in 2008. The baseline questionnaire included some questions on retrospective schooling history. Branson, Lam, and Zuze (2012) are able to match approximately 90% of enrolled youths to schools in the Department of Education's public database.

¹⁶One South African rand was equal to approximately \$0.145 in 2007 without adjusting for purchasing power parity.

the arc price elasticity of enrollment with respect to the value of fees

$$\begin{aligned}
\epsilon_2 &= \frac{\Delta \text{Enrollment rate}}{\Delta \text{Fee amount}} \times \frac{\text{Fee amount}}{\text{Enrollment rate}} \\
&= -\frac{0.0083}{237} \times \frac{0.6 \times 301 + 0.4 \times 64}{0.92} \\
&= -0.008
\end{aligned}$$

for the intention to treat estimate and - 0.009 for the instrumental variables estimate. The price elasticity of enrollment with respect to the total pecuniary cost of enrollment $\tilde{\epsilon}_2$ will be weakly larger. If, for example, fees are only half of the total cost of enrollment, then $\tilde{\epsilon}_2 = -0.023$ (ITT) or -0.026 (IV).¹⁷

The discussion above implies that fee elimination induced less than four additional students per school to enroll, increasing the baseline enrollment level by 1% and the baseline enrollment rate by less than one percentage point. Informal calculations suggest elasticities with respect to any fee or to the value of the fee no larger than -0.01. By any measure, this implies that demand for enrollment in the treated schools was highly price insensitive. Demand for broader measures of school participation that also take into account attendance by marginal students may be even lower. The remainder of this section shows that the result is robust to a variety of alternative estimation strategies. Section 2.5 explores potential explanations for this result.

Robustness checks: The results reported above are valid estimates of the causal effect of the fee elimination intervention on enrollment only if fee-charging and fee-eliminating schools would have experienced the same change in enrollment from 2005/6 to 2007/8 in the absence of the intervention. This assumption may fail if the two groups of schools have different observed or unobserved characteristics. To address this possibility, I estimate equations (2.1) – (2.3) using the weights in equation (2.4) to equalize the distribution of observed baseline school characteristics. Column 7 of table 2.1 shows that the difference in fee elimination rates between high and low poverty schools is unaffected by reweighting. However, the enrollment effect is halved from 3.6 students per school to 1.7 students per

¹⁷To derive this result, define E , F , and O as respectively the enrollment rate, the cost of fees and the cost of other education inputs (transport, uniforms, etc.). Then $\tilde{\epsilon}_2 = \frac{\Delta E}{\Delta(F+O)} \times \frac{F+O}{E}$. If O is unaffected by fee elimination, then $\tilde{\epsilon}_2 = \frac{\Delta E}{\Delta F} \times \frac{F}{E} + \frac{\Delta E}{\Delta F} \times \frac{O}{E} = \epsilon_2 \times (1 + \frac{O}{F})$.

school (95% confidence interval -1.5 to 4.9). This implies even more price insensitive demand than in the unweighted results. I also transform equations (2.3) – (2.3) from level to first difference specifications and include the vector of baseline school characteristics in the regression. The results, shown in columns 4 – 6 of table 2.1, are essentially identical to those from the unweighted regression. The “doubly-robust” analysis that uses regression and reweighting generates results very similar to those using just reweighting. All of these results suggest a very high rate of compliance with the intervention by schools and small effects on enrollment.¹⁸

An alternative approach to robustness is to estimate treatment effects for schools in the neighborhood of the poverty score cutoff between intention to treat and control schools. Figures 2.4 and 2.5 show estimates of the regression discontinuity models in equations (2.5) and (2.6) respectively. The estimates are obtain using local linear regression with bandwidth chosen to minimize the mean squared error of the treatment effect, following Imbens and Kalyanaraman (2009). The first figure shows that the probability of fee elimination rises from approximately 8% for schools just less poor than the cutoff to approximately 86% for schools just more poor than the cutoff. The difference of 76 percentage points is smaller than the difference between all intention to treat and control schools but is still substantial. The second figure shows that enrollment increases from 2005/6 to 2007/8 by 2.8 students more in intention to treat than control schools (confidence interval -0.7 to 6.3). This is slightly smaller than the difference for all schools, though the estimates are not statistically distinguishable at conventional significance levels. I also implement the regression discontinuity using global linear, quadratic and cubic models, and comparing means in the 50%, 25%, and 10% of the sample closest to the cutoff. The point estimates for the change in enrollment are somewhat sensitive to specification but are all between 0.8 and 4 students. This reinforces the earlier result that fee elimination had a relatively small effect on student enrollment.

As a final robustness check, I consider a longer time series of enrollment that includes four years of pre-treatment data. Figure 2.6 shows the raw enrollment data for intention to treat

¹⁸The difference between the weighted and unweighted results reflects heterogeneous enrollment trends over some observed characteristics, discussed in section 2.5 below. The unweighted estimators, with or without regression adjustment, are not consistent estimators of the relevant treatment effect in the presence of this form of heterogeneity. The identifying assumptions upon which the unadjusted and regression-adjusted models are based are subtly different, as discussed in Imbens and Wooldridge (2009).

and control schools. The two groups have very different pre-treatment levels of enrollment but there is little evidence of differential trends. The enrollment time series for the reweighted control schools looks almost identical to the treated schools. The graph supports the identifying assumption of parallel pre-treatment trends. I use a more formal test of this idea from Heckman and Hotz (1989) by constructing a “difference-in-second-differences” estimator. This equals the conventional difference-in-differences estimator (2007/8 versus 2005/6 enrollment in intention to treat and control schools) minus the pre-treatment difference-in-differences estimator (2005/6 versus 2003/4 enrollment in intention to treat and control schools). This is a consistent estimator of the treatment effect of interest under the weaker assumption that the two groups have linear but potentially non-parallel trends and are subject to common shocks. The estimated values of this parameter with and without reweighting are respectively 5.9 and 3.1 students (standard errors 4.7 and 4.2), compared to the standard estimate of 3.6 and 1.7 students (s.e. 1.8 and 1.6). This implies that pre-treatment enrollment in intention to treat schools was declining slightly relative to control schools. However, it still implies a small overall effect on enrollment.

2.5 Explaining the price-insensitive demand

This section builds off the conceptual framework developed in section 2.2 to explore possible reasons for the small enrollment effects that I find. The framework suggests that this may be due to (1) high baseline enrollment or “ceiling effects,” (2) low baseline fees, (3) inelastic demand, (4) binding capacity constraints, (5) credit constraints, or (6) negative effects on the valuation of enrollment. I do not directly observe baseline fees and so cannot test the second explanation. I explore the first, fourth and fifth explanations in this section and find that they cannot fully account for the small effects. The sixth explanation is left to the next section. The analysis as a whole suggests an important role for the residual explanation, inelastic demand.

Treatment effects by poverty level: The analysis to date has concentrated on quintile 2 schools and omitted the quintile 1 schools in poorer neighborhoods that also eliminated fees. Given the positive correlation between household income and enrollment found in survey data, baseline enrollment rates are likely to be lower in quintile 1 than quintile 2 schools.

There is thus less scope for ceiling effects on enrollment. I estimate quintile-specific versions of the difference-in-differences model in equations (2.1) – (2.3), with and without reweighting.¹⁹ Table 2.2 column 1 reports that the treatment compliance rates are very high: 99 and 95% for quintile 1 and 2 schools respectively. These are unaffected by reweighting. Enrollment increased from 2005/6 to 2007/8 in quintile 1 schools by 7.7 students (95% confidence interval 4.3 to 11 students), which represents a 2% increase in baseline enrollment. The equivalent change in quintile 2 schools is 3.4 students (confidence interval -0.1 to 7 students) or 0.9% of baseline enrollment. Combining these estimates using instrumental variables shows that fee elimination increased enrollment by an average of 6.1 students across all fee eliminating schools. As in the previous section, reweighting substantially reduces the magnitude of the treatment effects.

The significantly larger change in enrollment in quintile 1 schools may be due to ceiling effects on enrollment in quintile 2 schools. It may also reflect more elastic demand in quintile 1 schools.²⁰ This result is unlikely to be generated by differences in baseline school fees. Fees were higher in quintile 2 than quintile 1 schools (Borkum, 2011), implying larger treatment effects in the former unless students are credit constrained. I discuss the possibility of credit constraints in more detail below.

Grade-specific treatment effects: Baseline enrollment rates calculated from the General Household Survey are substantially higher for primary school-aged youths (96%) than secondary school-aged youths (85%). Even if these overstate the relevant enrollment rates due to grade repetition, they suggest that ceiling effects on enrollment are substantially more likely in primary than secondary school. I therefore estimate treatment effects on enrollment in each grade from 0 to 12 using basic and reweighted difference-in-differences models. The results are shown in figure 2.7. Panels A and C show that there are near-zero enrollment

¹⁹Reweighting quintile 3 schools to “look like” quintile 2 schools was successful in the sense that the mean values of the treatment and reweighted control schools’ baseline characteristics are neither substantively nor significantly different. Applying the same procedure to compare quintile 1 and 3 schools was less successful. The χ^2 test statistic for the null hypothesis that the means of all 28 baseline characteristics are equal across quintile 1 and reweighted quintile 3 schools is 54.3 (p -value 0.002). It is only possible to reduce this difference by using substantially higher order polynomial expressions in the reweighting functions in equation (2.4). The estimated treatment effects in quintile 1 should therefore be interpreted with a degree of caution.

²⁰These explanations are not entirely independent. A highly concave demand curve will result in near universal baseline enrollment and will be inelastic in the range affected by fee elimination.

changes in quintile 1 and quintile 2 primary schools respectively. The treatment effects are largest in the early grades of secondary school (8-10) and largely die out by grades 11 and 12. The reweighted estimates in panels B and D have a similar pattern though the point estimates are considerably smaller.

What do these effect sizes imply for enrollment rates? Averaging across quintile 1 and 2 schools, enrollment increases by 8.9 students in each of grades 8 and 9 and by 0.4 students in each grades 10, 11 and 12. These imply respective increases of 14.8% and 0.4% of the baseline enrollment level.²¹ The 2005 and 2006 General Household Surveys report enrollment rates of 87% for ages 14-15 and 83% for ages 16-18. Matching these age and grade brackets implies that eliminating fees increased age 14-15 enrollment from 87% to almost 100% and left age 16-18 enrollment almost unchanged at 83%.²²

Capacity constraints: Some schools may have a binding upper limit on the number of students they can accommodate, due to limited personnel, classroom space or other physical facilities. This would lead them to deny enrollment to students who would otherwise be induced to enroll by the policy change. The legal status of such denials is subject to an ongoing court challenge but anecdotal reports suggest that it is uncommon. I formally test for the existence of capacity constraints in two steps. I first calculate the maximum enrollment in each school between 2003 and 2006. This is one measure of each school's maximum capacity. I then compare the frequency with which 2007 or 2008 enrollment exceeds the previous maximum by quintile. This occurs in 26% of quintile 3 (control), 26% of quintile 2 schools, and 33% of quintile 1 schools. A substantial fraction of schools are thus able to accommodate additional students and this proportion is higher in fee-eliminating than fee-charging schools. Furthermore, the change in enrollment from 2006 to 2007 is smaller than at least one previous annual change in enrollment (2003 to 2004, 2004 to 2005 or 2005 to 2006) in over half of the sample. Hence, many schools are able to accommodate larger increases in enrollment than they experience when fees are eliminated. The results

²¹The difference between the level and percentage measures arise because the average school contains 53 students in each of grades 8 and 9, compared to 112 students in each of grades 10, 11 and 12.

²²Matching age and grade brackets assumes that no youths start school after the mandatory age, repeat grades or temporarily drop out of schooling. This assumption is clearly incorrect but unavoidable. In practice, many of the youths aged 16-18 will be enrolled in grades 8 and 9. This will bias the estimated treatment effect on the enrollment rate. The direction of the bias depends on the relative frequency of overage students in the inframarginal and marginal populations.

do not conclusively rule out capacity constraints but they do not appear to be of central importance.

Credit constraints: How important might credit constraints be in explaining the pattern of results discussed above? Section 2.2 noted that credit constraints might attenuate or augment the treatment effects of fee elimination. Edmonds (2006) reports some evidence of credit constraints to school enrollment in rural but not urban areas. Lam, Ardington, Branson, Goostrey, and Leibbrandt (2010) study a largely urban population and find little evidence of credit constraints to enrollment in tertiary education, which is typically much more expensive than primary or secondary enrollment. These results motivate estimating treatment effects separately for urban and rural schools.²³ Table 2.3 reports the results of this exercise. The compliance rates are equally high in both groups of schools. The enrollment effects are substantially larger in urban than rural areas, although the urban sample is smaller and the estimates less precise.²⁴

If credit constraints are indeed present in rural but not urban areas, these results suggest that there are a substantial number of rural students who would be induced to enroll by fee elimination if not for credit constraints. This is consistent with Edmonds' results, as the income shock he uses to test for credit constraints is substantially larger than the value of school fees. However, the results could also be due to larger baseline fees in urban areas or more elastic demand in urban areas. Even the larger effects in urban areas imply increases of less than 2% of baseline enrollment in quintile 1 and quintile 2 schools. Eliminating fees therefore leaves a large fraction of the population unenrolled in both rural and urban areas.

Taken together, these results demonstrate that the small enrollment effects are not explained by ceiling effects, capacity constraints or credit constraints. The next section explores whether they may be explained by a declining valuation of enrollment.

²³This includes all non-rural schools, urban and suburban.

²⁴The substantial difference between urban and rural schools' enrollment results explains the discrepancy between the reweighted and regression-adjusted results in table 2.1. The regression-adjusted estimates fail to take this heterogeneity into account.

2.6 Effects of fee elimination on school composition and outcomes

Fee elimination may change the resources available to schools and the composition of their student body. Compensating government transfers may be larger or smaller than the foregone fee revenue. Even if per-student revenues are unaffected, education inputs such as classrooms and teachers may adjust to changes in student numbers with a lag. Student expectations about changes in education resources may in turn affect their enrollment decisions and so attenuate or augment the enrollment response to fee elimination.

This section explores the equilibrium effects of fee elimination on education outcomes (measured by dropout rates, grade repetition rates, and class sizes) and the socio-economic profile of the enrolled students (measured by the proportion of students eligible for means-tested government social grants and the proportion who have had at least one parent die). These effects are equilibrium in that they are conditional on student enrollment decisions. They provide no direct information about the function mapping enrollment to education outcomes or *vice versa*. The section also explores the enrollment trends in control schools close to and far from treatment schools in order to address the possibility of spillovers.

Education outcomes: Table 2.4 column 2 shows fee elimination causes the dropout rate in fee-eliminating schools to fall from 2.8% to 2.1%. The rate of grade repetition also falls from 9.6% to 9.2% (column 4), though this effect is quite imprecisely estimated. Average class size increases from 39.3 to 39.8 students (column 6). In sum, fee elimination appears to have marginally increased enrollment and decreased per student education resources but reduced grade repetition and substantially reduced dropout. The dropout effect is most striking relative to its baseline level. This pattern would arise if marginal students are substantially less likely to drop out than their inframarginal peers or if fee elimination reduces credit constraint-induced dropout by inframarginal students. None of the other effects are large enough to suggest that education outcomes in schools are substantially affected by fee elimination. Future research will also examine whether high school completion, measured by passing a national graduation examination, is affected by fee elimination.

Socio-economic profile: The effect of fee elimination on the socio-economic profile of the enrolled student population is shown in columns 8 and 10 of table 2.4. The proportion

of students who have lost at least one parent to death falls from 20.7% to 19.5%²⁵ while the proportion of students eligible for means-tested government social grants stays constant. The socio-economic profile of the fee-eliminating schools appears relatively stable, suggesting that the marginal students are not substantially different on these dimensions to the inframarginal students. However, these data are collected as teachers' reports of students' self-reports and so are likely to be measured with substantial error. The effects should therefore be interpreted with caution.

Transfers between fee-eliminating and fee-charging schools: My estimation strategy assumes that the school fee elimination policy has no effect on enrollment levels at the control schools that continue to charge fees. This assumption may be violated if students who attended control schools before the policy change transferred to treatment schools after the policy change to take advantage of their lower cost. Such behavior would result in an upward bias in the estimated treatment effect of the fee elimination intervention.

I do not observe student-level data on transfers that would permit a direct test of this hypothesis. I therefore implement an indirect test that examines whether control schools that are geographically closer to treatment schools experience falls in enrollment from 2006 to 2007 relative to farther away control schools and relative to their own enrollment change in previous years. Figure 2.8 shows a local linear regression of the change in grade-level enrollment from 2006 to 2007 at control schools against the distance from the nearest treatment school offering the same grade.²⁶ Control schools nearer to treatment schools actually experience small gains in enrollment relative to control schools farther away. I cannot reject that this pattern of changes is identical to that observed between 2005 and 2006, before the fee elimination policy was implemented. The result is robust to restricting the sample to control schools within one half standard deviation of the cutoff. I interpret this as strong evidence against the spillover hypothesis.

I also estimate a linear regression of change in enrollment by grade at fee-charging schools from 2006 to 2007 on the same measure at the nearest fee-eliminating school. If the treatment effect is driven entirely by transfers, the slope coefficient should be approximately equal to

²⁵The effect on the proportion of students who have lost both parents to death falls by 0.2 percentage points from a baseline of 3.8%. The patterns for both definition of orphan are therefore consistent.

²⁶I construct this distance measure using GIS codes for every school in the sample.

-1. Instead, it equals 0.045 (standard error 0.016). This is not significantly different to the coefficient in the equivalent regression using changes from 2005 to 2006 (0.037, with standard error 0.08). This result is robust to weighting the regression by the inverse distance between treatment and control schools, to restricting the sample to control schools within one half standard deviation of the cutoff, and to excluding control schools that are more than 10 miles from the nearest treatment school. These results strongly suggest that the treatment effects are not driven by transfers between schools. However, I observe only net transfers and not gross inflow and outflow of students into each school. I cannot rule out the possibility that approximately equal numbers of students transfer in each direction between fee-charging and fee-eliminating schools.

One interpretation of these results is that geographically targeted variation in the prices of public services may be an effective alternative to individual-level means testing. The absence of spillovers in this context suggests that setting lower prices for public services in poor neighborhoods does not induce people from wealthier neighborhoods to adjust their behavior to take advantage of the lower cost services. Geographic targeting may be a desirable alternative to individual means-testing when the latter is expensive. However, caution should be exercised in generalizing this result to settings where use of public services is more price-sensitive or transport costs are lower.

2.7 Conclusion

Increasing participation in formal schooling is an important thrust of public policy throughout the developing world. A wide range of countries have employed a wide range of demand- and supply-side interventions to increase student enrollment and attendance. Empirical work in microeconomics has found evidence of positive relationships between formal schooling and individual earnings, health and future children's human capital accumulation, at least some of which seem to capture causal effects of education. The macroeconomic literature points to a potential role of formal schooling in explaining cross-country differences in per capita income and growth.

This paper contributes to this literature and policy by studying the effect of geographically targeted school fee eliminations on primary and secondary school enrollment in South

Africa. Fee-reducing and -eliminating interventions have become popular in developing countries over the past two decades. There is, however, relatively little empirical evidence on their effects. I find that eliminating fees to enroll in South African schools in high-poverty neighborhoods increased enrollment by a relatively small margin. My preferred estimates suggest that 1 to 3 additional students were induced to enroll in the average school. This increased the baseline enrollment level by 1% or less and the baseline enrollment rate by less than 1 percentage point. Back-of-the-envelope calculations suggest a price elasticity of enrollment in the neighborhood of -0.01. School-level composition, education outcomes and resources are also largely unaffected by fee elimination.

I explore a variety of explanations for this pattern. The results are not explained by pre-treatment time trends, pre-treatment differences between high- and low-poverty schools' characteristics, overstated enrollment levels, or transfers between fee-eliminating schools and fee charging schools. They are not fully explained by ceiling effects on enrollment, capacity constraints in schools or credit constraints in households.

I therefore conclude that demand for schooling is relatively price insensitive in the neighborhoods treated by the fee elimination intervention. This is more likely due to low returns to education for marginal students in these schools than to high labor market opportunity costs. A substantial body of prior research has documented low "quality" of education in high-poverty South African schools, measured by enrolled students' test scores and graduation rates. Particularly low returns to enrollment for marginal students would be unsurprising.

My results imply that reducing enrollment costs may have limited impact on school participation levels in settings where enrollment does not translate into substantial learning or grade progression. This emphasizes a potential complementarity between cost-reduction and quality-upgrading interventions and points to an important avenue for future research.

Table 2.1: Treatment effects of fee elimination on school-level enrollment

| Outcome | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) |
|---|---------------------|------------------|------------------|---------------------|-----------------|-----------------|---------------------|----------------|---------------------|---------------------|------------|----------------|
| Estimator | Fee Elimination OLS | Enroll OLS | Enroll IV | Fee Elimination OLS | Enroll OLS | Enroll IV | Fee Elimination OLS | Enroll OLS | Enroll IV | Fee Elimination OLS | Enroll OLS | Enroll IV |
| High poverty | 0.944*** (0.004) | | | 0.941*** (0.004) | 3.24* (1.84) | | 0.943*** (0.004) | | 0.944*** (0.004) | 1.51 (1.59) | | |
| High poverty \times post treatment | | 3.57** (1.81) | | | | | | 1.72 (1.63) | | | | |
| Eliminated fees | | | | | | 3.44* (1.95) | | | | | | 1.60 (1.68) |
| Eliminated fees \times post treatment | | | 3.78** (1.92) | | | | | | 1.82 (1.72) | | | |
| Regression | | | | \times | \times | \times | \times | \times | \times | \times | \times | \times |
| Reweighting | | | | | | | | | | | | |
| Adjusted R2 | 0.901 | 0.049 | 0.047 | 0.905 | 0.052 | 0.052 | 0.904 | 0.000 | \times | 0.908 | 0.047 | 0.047 |
| # clusters | 6181 | 6181 | 6181 | 6181 | 6181 | 6181 | 6181 | 6181 | 6181 | 6181 | 6181 | 6181 |
| # observations | 12362 | 24724 | 24724 | 12362 | 12362 | 12362 | 12362 | 24724 | 24724 | 12362 | 12362 | 12362 |

Notes: Columns 1, 4, 7 and 10 show the rates of fee elimination in high-poverty schools (quintile 2) relative to low-poverty schools (quintile 3). Columns 2, 5, 8 and 11 show the change in enrollment from 2005/6 to 2007/8 in high-poverty schools relative to low-poverty schools. Columns 3, 6, 9, and 12 show the change in enrollment from 2005/6 to 2007/8 in fee-eliminating schools relative to fee-charging schools with fee status instrumented by poverty level. Columns 1, 2, 3, 7, 8, and 9 are estimated in levels using data from both periods. Columns 4, 5, 6, 10, 11, and 12 are estimated in differences and control for baseline school characteristics using regression. Columns 7, 8, 9, 10, 11, and 12 are estimated using weighted least squares with weights assigned to low poverty schools to equate their distribution of baseline characteristics with those of the high poverty schools. Standard errors are shown in parentheses. In columns 1, 2, 3, 4, 5, and 6 the standard errors are estimated using a cluster-robust variance estimator allowing unrestricted error correlation within each school unit. In columns 7, 8, 9, 10, 11, and 12 the standard errors are constructed from 100 replications of a bootstrap algorithm that resamples school units and is stratified by treatment group. *, **, and *** denote significance at the 10, 5, and 1% levels respectively.

Table 2.2: Treatment effects of fee elimination on enrollment in high- and very-high poverty schools (quintiles 2 and 1)

| | (1) | (2) | (3) | (4) | (5) | (6) |
|--|----------------------|-------------------|-------------------|----------------------|-------------------|------------------|
| Outcome | Fee Eli- mination | Enroll | Enroll | Fee Eli- mination | Enroll | Enroll |
| Estimator | OLS | OLS | IV | OLS | OLS | IV |
| High poverty | 0.949*** (0.004) | | | 0.945*** (0.004) | | |
| Very high poverty | 0.995*** (0.001) | | | 0.995*** (0.002) | | |
| High poverty \times post-treatment | | 3.41* (1.81) | | | 0.67 (1.74) | |
| Very high poverty \times post-treatment | | 7.65*** (1.73) | | | 4.91*** (1.66) | |
| Eliminated fees \times post-treatment | | | 6.13*** (1.66) | | | 3.26** (1.58) |
| Reweighting | | | | \times | \times | \times |
| Adjusted R2 | 0.930 | 0.059 | 0.058 | 0.957 | 0.001 | 0.001 |
| # clusters | 10235 | 10235 | 10235 | 10235 | 10235 | 10235 |
| # observations | 20470 | 40940 | 40940 | 20470 | 40940 | 40940 |

Notes: Columns 1 and 4 show the rates of fee elimination in high- and very high-poverty schools (quintiles 2 and 1) relative to low-poverty schools (quintile 3). Columns 2 and 5 show the change in enrollment from 2005/6 to 2007/8 in high- and very high-poverty schools relative to low-poverty schools. Columns 3 and 6 show the change in enrollment from 2005/6 to 2007/8 in fee-eliminating schools relative to fee-charging schools with fee status instrumented by indicators for high- and very high-poverty schools. Columns 4, 5, and 6 are estimated using weighted least squares with weights assigned to low poverty schools to equate their distribution of baseline characteristics with those of the high- and very high-poverty schools. Standard errors are shown in parentheses. In columns 1, 2 and 3 these are estimated using a cluster-robust variance estimator allowing unrestricted error correlation within each school unit. In columns 4, 5 and 6 these are constructed from 100 replications of a bootstrap algorithm that resamples school units and is stratified by treatment group. *, **, and *** denote significance at the 10, 5, and 1% levels respectively.

Table 2.3: Treatment effects of fee elimination on enrollment in rural and urban schools

| | (1) | (2) | (3) | (4) | (5) | (6) |
|--|----------------------|------------------|-------------------|----------------------|------------------|-----------------|
| Outcome | Fee Eli- mination | Enroll | Enroll | Fee Eli- mination | Enroll | Enroll |
| Estimator | OLS | OLS | IV | OLS | OLS | IV |
| Sample | Urban schools | | | Rural schools | | |
| High poverty | 0.979*** (0.006) | | | 0.944*** (0.005) | | |
| Very high poverty | 1.000*** (0.000) | | | 0.997*** (0.001) | | |
| High poverty \times post-treatment | | 9.18** (3.93) | | | -0.04 (1.93) | |
| Very high poverty \times post-treatment | | 7.50** (3.83) | | | 4.75** (1.86) | |
| Eliminated fees \times post-treatment | | | 8.40*** (3.27) | | | 3.12* (1.79) |
| Reweighting | | | | \times | \times | \times |
| Adjusted R2 | 0.982 | 0.050 | 0.049 | 0.902 | 0.016 | 0.017 |
| # clusters | 2791 | 2791 | 2791 | 7293 | 7293 | 7293 |
| # observations | 5582 | 11164 | 11164 | 20470 | 40940 | 40940 |

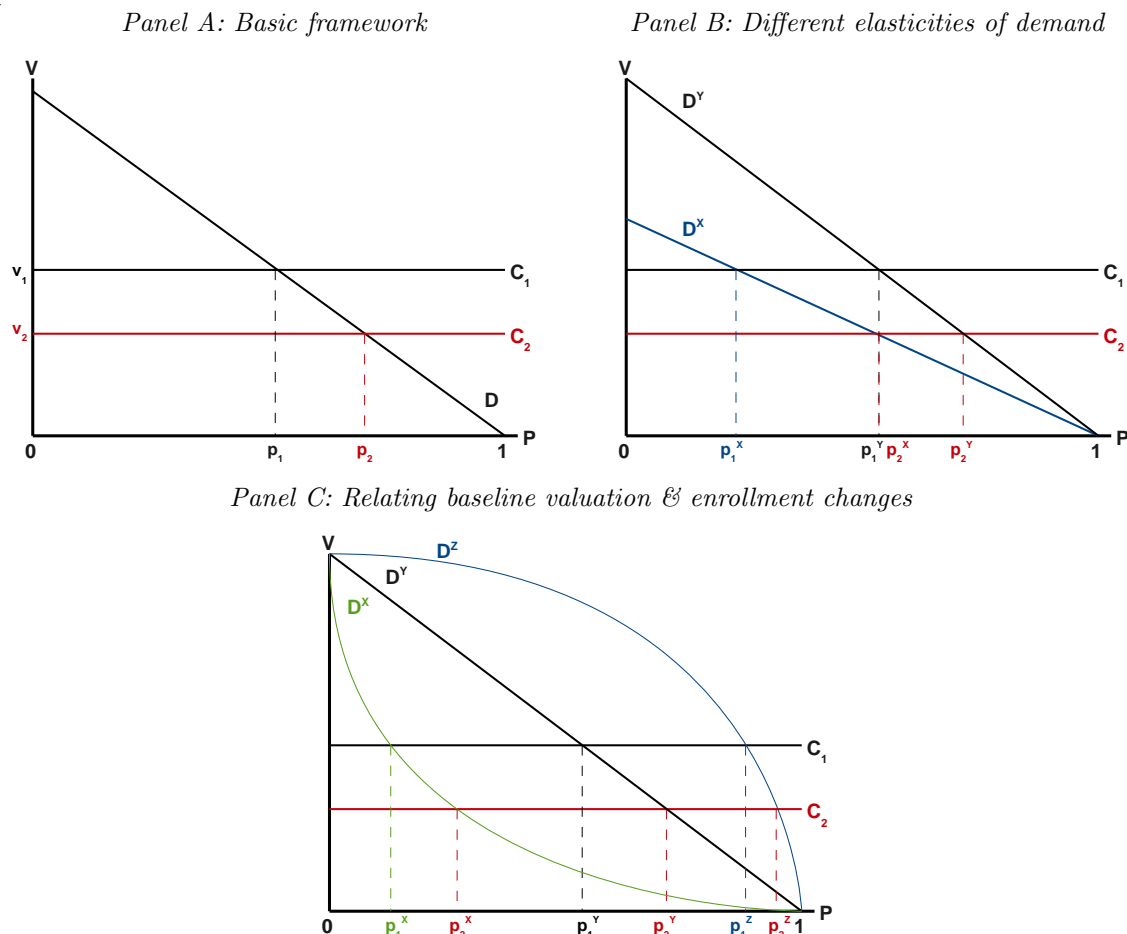
Notes: Columns 1 and 4 show the rates of fee elimination in high- and very high-poverty schools (quintiles 2 and 1) relative to low-poverty schools (quintile 3). Columns 2 and 5 show the change in enrollment from 2005/6 to 2007/8 in high- and very high-poverty schools relative to low-poverty schools. Columns 3 and 6 show the change in enrollment from 2005/6 to 2007/8 in fee-eliminating schools relative to fee-charging schools with fee status instrumented by indicators for high- and very high-poverty schools. Standard errors are shown in parentheses and are estimated using a cluster-robust variance estimator allowing unrestricted error correlation within each school unit. *, **, and *** denote significance at the 10, 5, and 1% levels respectively.

Table 2.4: Treatment effects of fee elimination on measures of school resources, grade progress, and composition

| Outcome | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
|---|--------------------|--------------------|-------------------|-----------------|-------------------|------------------|------------------|-----------------|--------------------|--------------------|
| Estimator | OLS | IV | OLS | IV | OLS | IV | OLS | IV | OLS | IV |
| High poverty \times post-treatment | -0.28** (0.11) | | -0.44** (0.22) | | 0.89*** (0.30) | | 1.66** (0.74) | | -0.98*** (0.32) | |
| Very high poverty \times post-treatment | -0.96*** (0.10) | | -0.14 (0.20) | | 0.28 (0.25) | | -1.10 (0.72) | | -1.18*** (0.31) | |
| Eliminated fees \times post-treatment | | -0.72*** (0.09) | | -0.26 (0.19) | | 0.50** (0.23) | | -0.03 (0.65) | | -1.13*** (0.28) |
| Pre-treatment mean | | 2.78 | | 9.63 | | 39.31 | | 37.31 | | 20.66 |
| Adjusted R2 | 0.003 | 0.002 | 0.003 | 0.002 | 0.006 | 0.006 | 0.003 | 0.003 | 0.002 | 0.002 |
| # clusters | 9857 | 9857 | 9856 | 9856 | 10147 | 10147 | 10235 | 40929 | 10131 | 36282 |
| # observations | 33399 | 33399 | 33396 | 33396 | 36986 | 36986 | 10235 | 40929 | 10131 | 36282 |

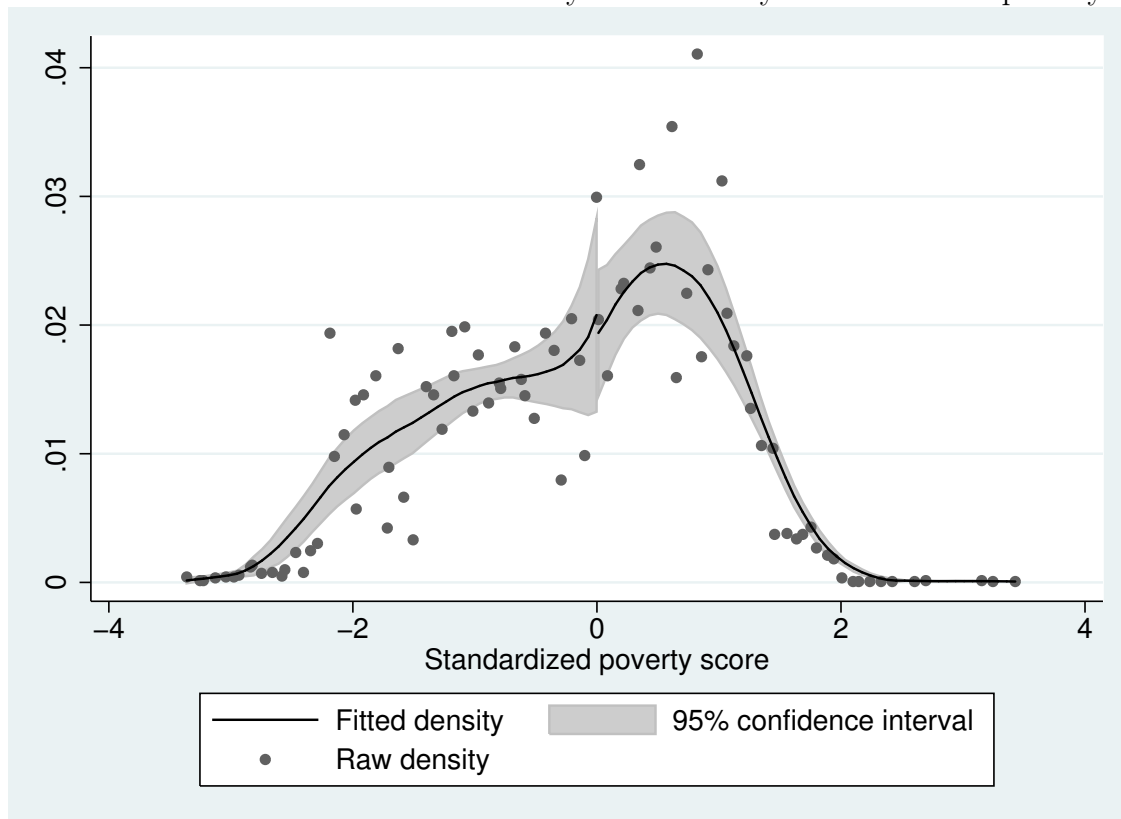
Notes: Columns 1, 3, 5, 7 and 9 show the change in the relevant outcome from 2005/6 to 2007/8 in high-and very high-poverty schools relative to low-poverty schools. Columns 2, 4, 6, 8 and 10 show the change in the relevant outcome from 2005/6 to 2007/8 in fee-eliminating schools relative to fee-charging schools, with fee status instrumented by indicators for high- and very high-poverty schools. Standard errors are shown in parentheses and are estimated using a cluster-robust variance estimator allowing unrestricted error correlation within each school unit. *, **, and *** denote significance at the 10, 5, and 1% levels respectively.

Figure 2.1: Conceptual framework showing the effect of fee elimination $C_1 \rightarrow C_2$ on the proportion of students enrolled P



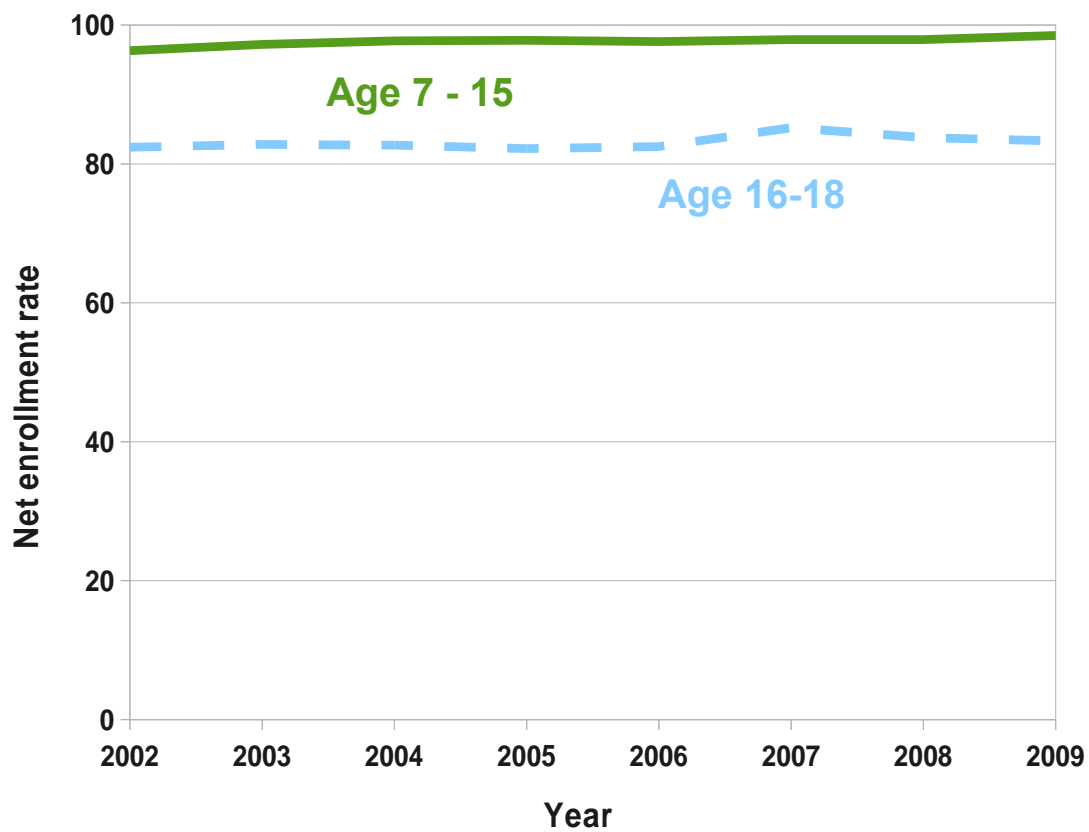
Notes: This figure illustrates the conceptual framework used in the paper. The proportion of students enrolled P is increased by eliminating school fees (panel A), the magnitude of the effect is larger when demand is more elastic (panel B), and the magnitude of the effect need not be correlated with mean baseline valuation (panel C).

Figure 2.2: Falsification test for a discontinuity in the density of standardized poverty scores



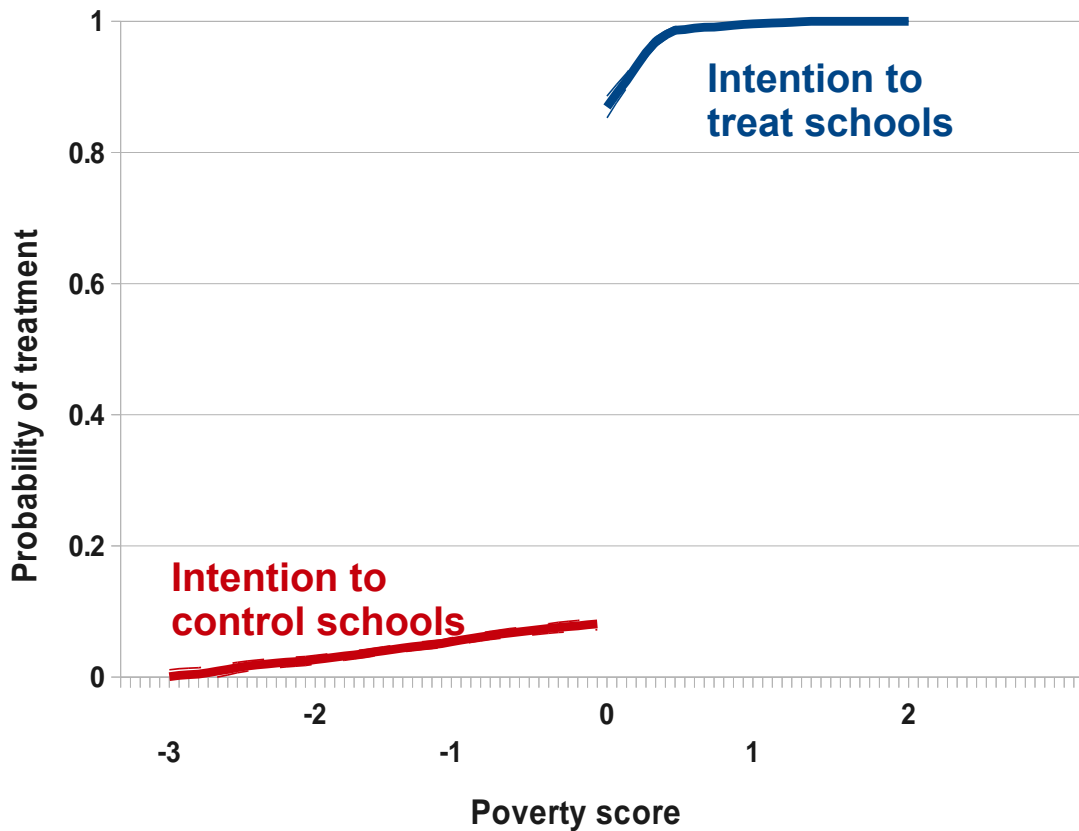
Notes: This figure estimates the density of poverty scores for schools in low- and high-poverty neighborhoods (left- and right-hand sections of the figure respectively). The difference between the densities evaluated at the cutoff between low- and high-poverty schools is -0.002 and is not statistically significant. This provides reassuring evidence that there was not systematic manipulation of poverty scores in order to control schools' assignment to (intended) treatment status. The densities are estimated separately on either side of the cutoff using local linear regression with a plug-in bandwidth selection. The results are robust to alternative bandwidth choices.

Figure 2.3: Time trend in net enrollment rate



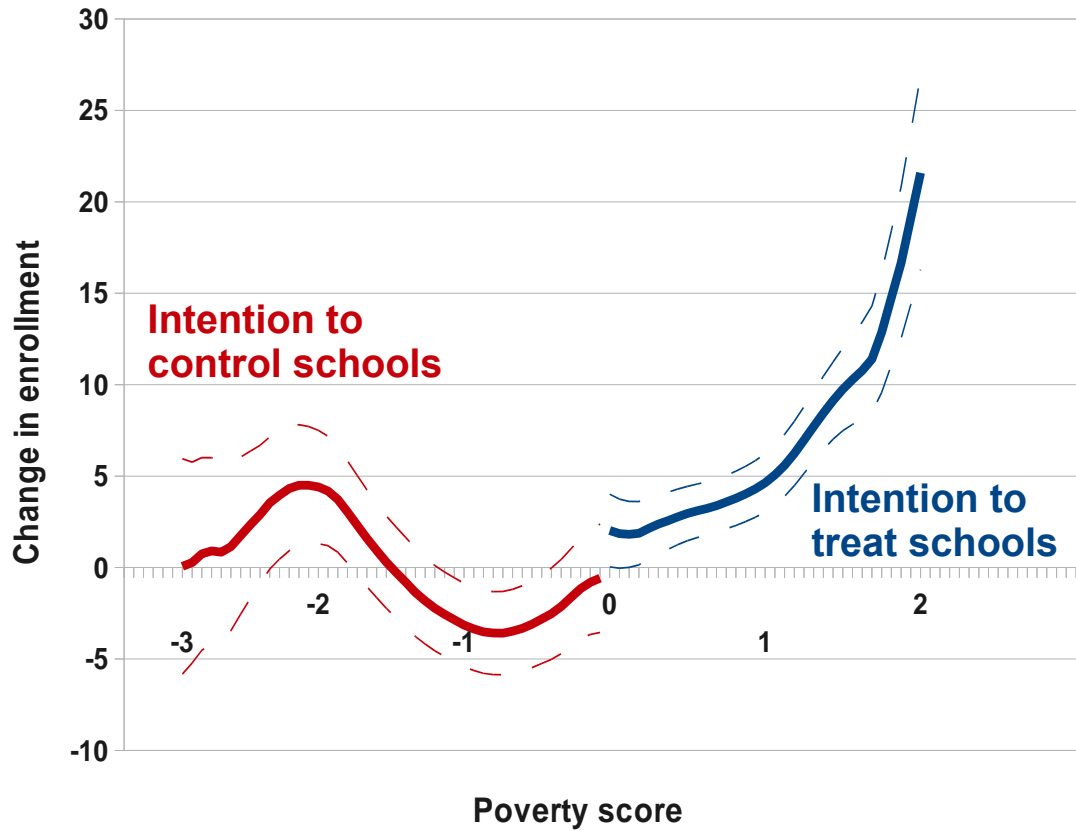
Notes: This figure shows the time series of the net enrollment rate. Data are calculated from the General Household Survey using appropriate sampling weights.

Figure 2.4: Probability that schools eliminate fees, by poverty score



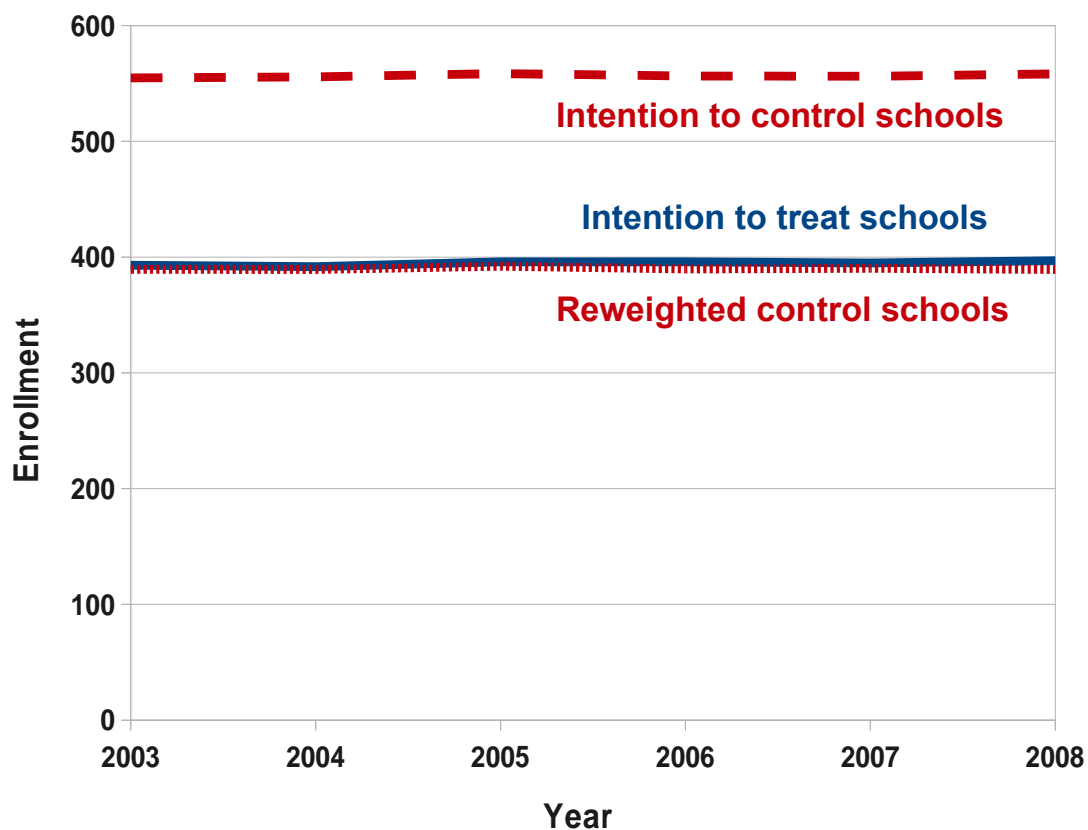
Notes: This figure shows the fitted probability that schools eliminate fees by their assigned poverty score. The fitted curves and 95% confidence intervals are from local linear regressions estimated separately on either side of the cutoff, with bandwidth choices following Imbens and Kalyanaraman (2009). The estimated difference between the curves at the threshold value is 76 percentage points with a standard error of 1 percentage point.

Figure 2.5: Change in enrollment from 2005/6 to 2007/8, by poverty score



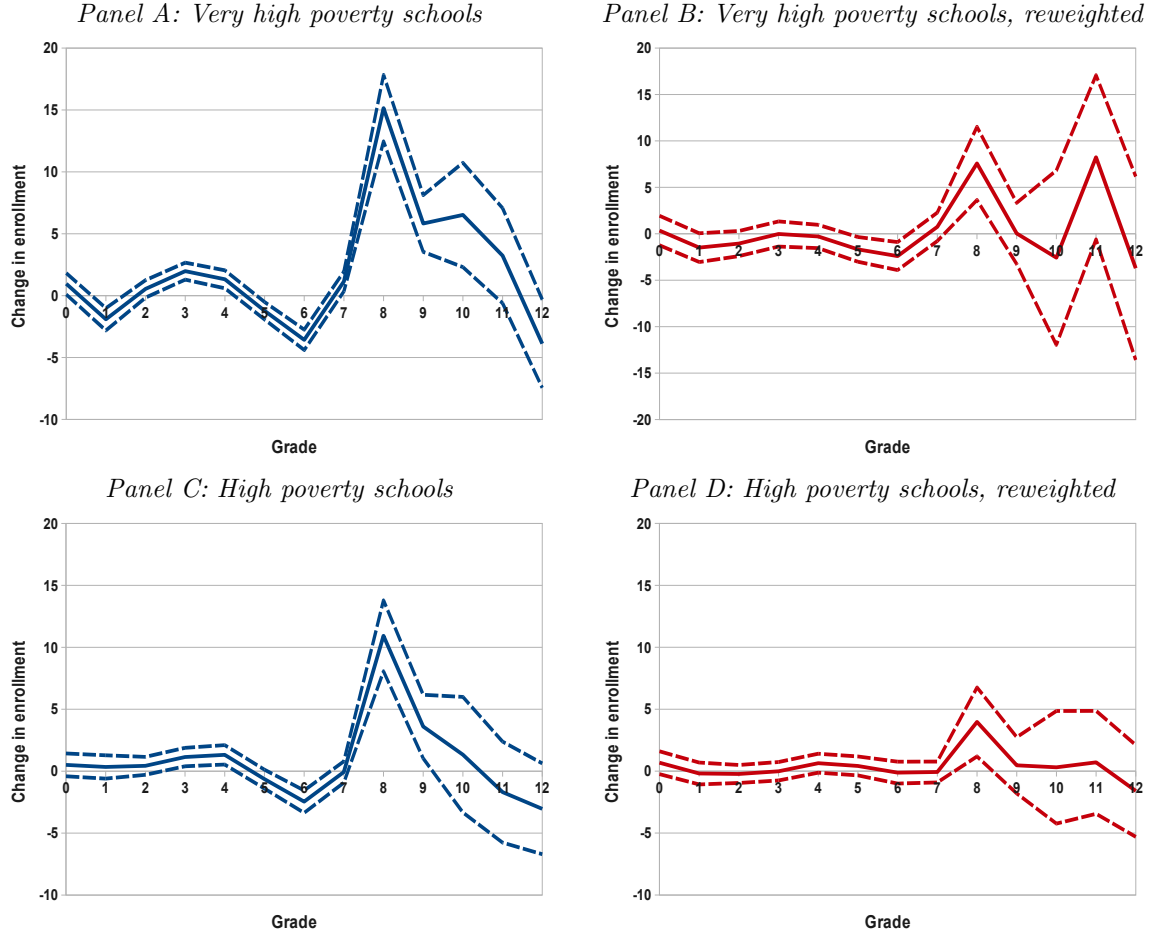
Notes: This figure shows the fitted change in school enrollment from 2005/6 to 2007/8 by schools' assigned poverty scores. The fitted curves and 95% confidence intervals are from local linear regressions estimated separately on either side of the cutoff, with bandwidth choices following Imbens and Kalyanaraman (2009). The estimated difference between the curves at the threshold value is 2.8 students with a standard error of 1.8 students.

Figure 2.6: Level of enrollment from 2003 to 2008 for intention to treat schools, control schools, and reweighted control schools



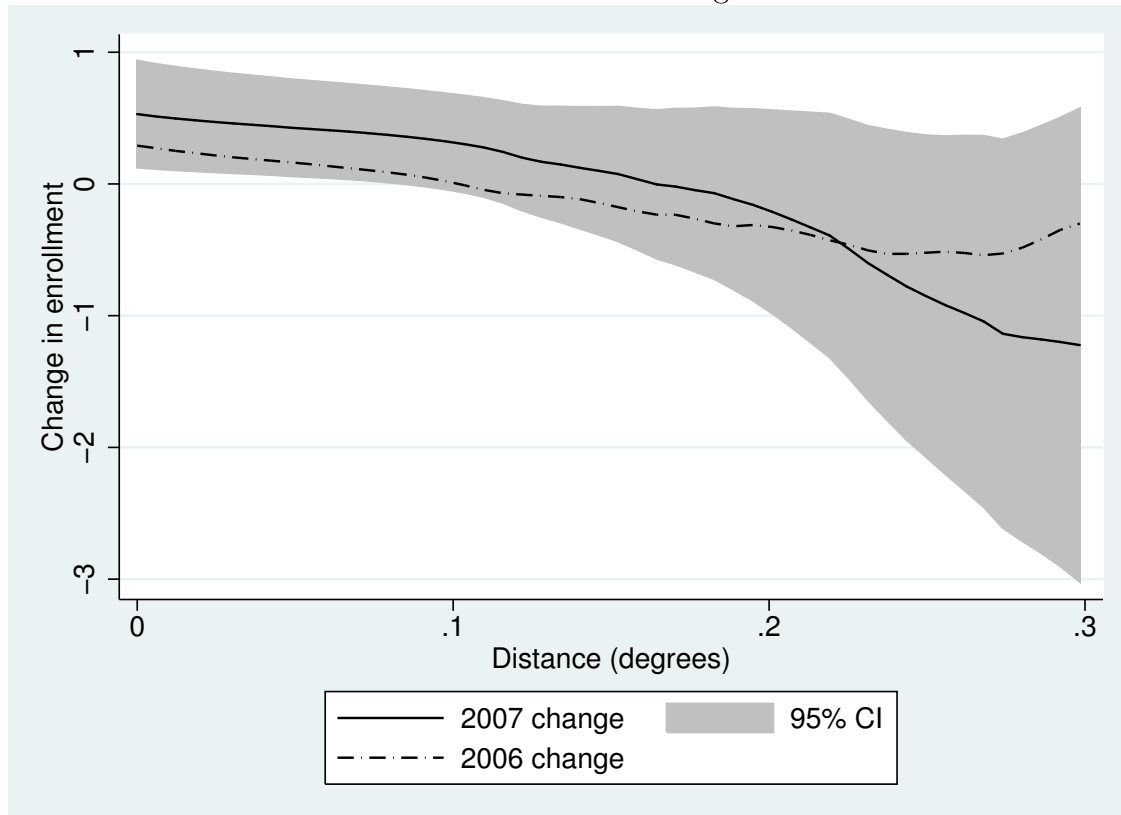
Notes: This figure shows the level of enrollment for different groups of schools in each year from 2003 to 2008. The solid blue line shows intention to treat (quintile 2) schools, the dashed red line shows intention to control (quintile 3) schools, and the dotted red line shows the latter group of schools after reweighting them to have the same distribution of baseline observed characteristics as the intention to treat schools.

Figure 2.7: Treatment effects of fee elimination on enrollment by grade



Notes: This figure shows grade-specific treatment effects of fee elimination on enrollment. Panels A and B show results for schools in very high poverty neighborhoods (quintile 1) while panels C and D show results for schools in less high poverty neighborhoods (quintile 2). Panels B and D show results after reweighting the control schools (quintile 3) to have the same distribution of baseline observed characteristics. These are intention-to-treat effects. The compliance rate does not vary substantially by grade so the overall pattern of instrumental variables effects is similar. The 95% confidence intervals in panels A and C are obtained from a school-level cluster robust variance estimator. The intervals in panels B and D are obtained from 100 replications of a bootstrap that resamples schools and stratifies by quintile.

Figure 2.8: Changes in enrollment from 2005 to 2006 and 2006 to 2007 for fee-charging schools located at different distances from fee-eliminating schools



Notes: This figure uses local linear regression to estimate the relationship between the change in enrollment from 2006 to 2007 and proximity to a fee-eliminating school (solid line) for the sample of fee-charging schools. It estimates the same relationship for the change from 2005 to 2006 (dashed line). A strong negative slope for the former relationship relative to the latter relationship would suggest that fee-charging schools near fee-eliminating schools are losing students to those treated schools. The results strongly reject this hypothesis and are robust to alternative bandwidth choices.

Chapter 3

Mobility Treatment Effects: Identification, Estimation, and Application

3.1 Introduction

The measurement and analysis of intertemporal economic mobility is an important subject in both academic research and policy analysis. Intergenerational economic mobility has featured prominently in the practice and study of politics. Economists and other social scientists have studied how educational attainment, income, and occupation changes through time and between generations. Developing satisfactory measures of economic mobility has been a central theme of this research agenda.

In this paper I demonstrate how to interpret intertemporal economic mobility measures in a treatment effects or decomposition framework. I consider the problem of comparing levels of economic mobility across two groups. I develop a set of sufficient conditions under which this difference can be causally attributed to group membership. Equivalently, I develop a set of sufficient conditions for identification of *mobility treatment effects*. My framework also demonstrates how mobility differences across groups can be decomposed into components explained and not explained by differences in observed characteristics. This decomposition exercise is identified under the same set of conditions required for causal analysis, though the economic interpretation of the conditions may be different.

Consider a simple example to illustrate this paper's contribution. Assume that income Y is observed for two groups $D \in \{0, 1\}$ in two periods $T \in \{0, 1\}$. In period 0, half the members of each group have income below a time-invariant poverty line \bar{Y} . The income

dynamics in the two groups are summarized by the transition matrices

$$\begin{aligned} M^0 &= \begin{pmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} \end{pmatrix} \\ M^1 &= \begin{pmatrix} \frac{1}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{1}{3} \end{pmatrix} \end{aligned} \tag{3.1}$$

for groups $D = 0$ and $D = 1$ respectively. The first matrix indicates that of individuals below the poverty line in period 0, two thirds remain below the line in period 1 (cell m_{11}^0) and one third move above the poverty line (cell m_{12}^0). Of individuals above the poverty line in period 0, one third move below the poverty line in period 1 (cell m_{21}^0) and two thirds remain above the poverty line (cell m_{22}^0). In contrast, the second matrix indicates that two thirds of individuals below the poverty line in period 0 move above it in period 1 (cell m_{12}^1) and *vice versa* for individuals above the poverty line in period 0 (cell m_{21}^1).

The simplest way to summarize the information in these matrices is to note that the probability that an individual will cross the poverty line is $1/3$ in group 0 and $2/3$ in group 1. Group 1 therefore has a higher level of intertemporal economic mobility by this metric. My contribution is to specify sufficient conditions for this difference in mobility to be assigned a causal interpretation as a mobility treatment effect. This contribution also clarifies under which assumptions the difference in mobility can be attributed to cross-group differences in education, demographic factors, etc.

Why is the difference in economic mobility of interest to economists or to policymakers? Note that the difference between the groups would not be captured by static comparisons in each period. The poverty rate is 50% in each group in each period. The mobility measure therefore conveys new information that may be of substantive importance for several reasons. First, welfare depends on individuals' stream of intertemporal consumption in many models of habit formation and lifecycle consumption (Atkinson, 1983). Second, mobility measures are well suited to analyzing how individuals' ranks in a distribution change through time or how individuals' ranks depend on those of their parents. These ranks, and their changes through time, play a central role in the political economy literature (Galor and Tsiddon,

1997) and in models where welfare depends on the formation of reference points (Becker and Murphy, 2000). Third, the incidence of some public policies depends on intertemporal economic mobility. Some means-tested social programs have maximum lifetime eligibility conditions, which bind for individuals in long-term poverty but not individuals in transient poverty. Relatedly, the incidence of some public policies such as access to capacity constrained education institutions may depend on individual ranks with respect an economic outcome (Cullen, Long, and Reback, 2013; Ozier, 2011). Understanding the intertemporal path of individual ranks becomes important in such settings.

The framework I develop in this paper will be useful in three settings. First, mobility measures provide a useful complement to existing treatment effects estimands, whether used in program evaluation or in more general economic analysis. Mobility measures provide information that is not captured by standard estimands such as average and quantile treatment effects. This information may allow more comprehensive evaluation of the welfare effects of treatments or facilitate differentiation between different theoretical models. Second, economic mobility is often compared across substantially different groups (Solon, 2002). My framework clarifies under what conditions these mobility differences can be interpreted causally and/or be decomposed into the relative contributions of different group characteristics.

Third, mobility measures may also be useful in the analysis of non-cardinal data. Standardized test scores, for example, are often normed with reference to the test-taking population and are not strictly cardinal. This makes comparison across settings difficult, such as interpreting level differences in normed test scores between groups arising from an experimental intervention. Normed data can, however, be used to explore how certain measures of intertemporal mobility differ across arms of the intervention. For example, the number of individuals scoring below the median in period 0 who move above the median in period 1 can be analyzed independent of the outcome scaling. This is necessarily a zero-sum measure (the same number of individuals must move in the other direction) but it still provides information about whether different arms of the intervention increase or reduce mobility.

Organization of the paper: Section 3.2 develops the key results in the paper for identification of mobility treatment effects. The key conditions are that group assignment is

uncorrelated with unobserved individual characteristics, conditional on baseline outcomes, and that the baseline outcomes between the two groups have common support. I show that these conditions are satisfied by randomized group assignment, and by some other research designs. The conditions are similar to but weaker than the “selection on observed variables” assumption commonly employed in applied research. I also note that with more than two time periods, the key identification assumption is testable. Section 3.3 discusses estimation and inference for mobility treatment effects. The key results are already established in the econometrics literature so I merely note how they apply in this context. I discuss an application of the ideas in this paper in section 3.4. I estimate MTEs for an education experiment in Kenya that studied the effects of classroom tracking relative to random assignment on test scores. I find that tracking marginally increased the level of mobility and show that this provides additional support for the authors’ interpretation of the experiment. Section 3.5 concludes.

Related literature: This paper builds off three literatures. The first concerns the definition and measurement of mobility. Early mobility measurement started with transition matrices like those in equation (3.1). Mobility measures were defined as scalar summary statistics over transition matrices (Prais, 1955; Shorrocks, 1978b; Sommers and Conlisk, 1979). Subsequent work focused on deriving mobility criteria from dynamic social welfare functions, based on changes in individual ranks or levels through time (Atkinson, 1983; Fields and Ok, 1996; Shorrocks, 1978a). Many of these mobility measures are based on stochastic dominance criteria over distributions. These are less susceptible to counter-intuitive rankings of special cases than the transition matrix summary statistics but do not necessarily yield complete orderings over all distributions (Bartholomew, 1982).

The second literature concentrates on estimating mobility measures and performing inference on these estimates. The estimation problem is typically straightforward: almost all mobility measures are continuous functions of moments and order statistics and so can be consistently estimated by sample analogues. Inference has proved slightly more problematic. Several papers establish analytical variance estimators for a wide range of mobility measures (Schluter, 1998; Formby, Smith, and Zheng, 2004) but these assume independently and identically distributed data and so are not appropriate for clustered or weighted data. Biewen

(2002) establishes the validity of the bootstrap for most mobility measures but Davidson and Flachaire (2007) argue that the bootstrap-based tests may be very sensitive to the distribution of the upper tail of the data and this leads to confidence intervals with incorrect empirical coverage. This criticism does not apply to rank-based mobility measures, which monotonically transform data into a uniform scale.

The third literature consists of empirical applications of mobility measurement. Most studies focus on intergenerational earnings, education, or occupation mobility across generations, or intertemporal income mobility for individuals. See Solon (1999) for a review of this empirical literature in economics.

3.2 Identification

This section of the paper lays out sufficient conditions for causal analysis of mobility measures. I use a potential outcomes framework throughout, with $D = 1$ and $D = 0$ denoting the treatment and control groups. Let Y^T denote pre- and post-treatment outcomes for $T = 0$ and $T = 1$ respectively, and let \tilde{Y} denote outcomes under treatment. Mobility measures are functions¹ of the joint distribution of pre- and post-treatment outcomes in either group

$$\mu \left(F_{Y^1, Y^0|D} (\cdot, \cdot) \right). \quad (3.2)$$

I define a *mobility treatment effect* (MTE) as the difference between the observed mobility in the treatment group and the counterfactual level of mobility that this group would have experienced in the absence of treatment²

$$\Delta^{\text{MTE}} = \underbrace{\mu \left(F_{\tilde{Y}^1, Y^0|D=1} (\cdot, \cdot) \right)}_{\text{Observed}} - \underbrace{\mu \left(F_{Y^1, Y^0|D=1} (\cdot, \cdot) \right)}_{\text{Unobserved}} \quad (3.3)$$

¹Measures may be scalars or vectors. In the motivating example, the transition matrices M^0 and M^1 are vector-valued mobility measures while the proportion of individuals crossing the poverty line is a scalar mobility measure. The identification results also apply to functional mobility measures, which I do not discuss here.

²I limit this discussion to identification of *mobility treatment effects on the treated*. The identification conditions for *mobility treatment effects on the untreated* are analogous. Combining the two identification conditions yields the set of mobility treatment effects. See Heckman and Robb (1985) for a discussion of this distinction in the context of average treatment effects and Athey and Imbens (2006) for a similar argument regarding combining treatment effects on the treated and untreated.

The key contribution of the paper is develop a set of sufficient identifying assumptions under which the unobserved counterfactual level of mobility in the treatment group can be replaced by some transformation of the observed level of mobility in the control group $\mu(F_{Y^1, Y^0|D=0}(\cdot, \cdot))$.

Theorem 1. *If*

A1 $Y^1 \perp D|Y^0$ and

A2 $\text{support}(Y^0|D=1) \subseteq \text{support}(Y^0|D=0)$

then the unobserved joint distribution of pre- and post-treatment outcomes in the treatment group $F_{Y^1, Y^0|D=1}^{CF}(\cdot, \cdot)$ equals the rescaled observed joint distribution of pre- and post-treatment outcomes in the control group

$$F_{Y^1, Y^0|D=0}(y^1, y^0) \times \frac{F_{Y^0|D=1}(y^0)}{F_{Y^0|D=0}(y^0)}. \quad (3.4)$$

Proof. Using the definition of a conditional probability distribution,

$$\begin{aligned} F_{Y^1, Y^0|D=1}^{CF}(y^1, y^0) &= Pr(Y^1 \leq y^1, Y^0 \leq y^0|D=1) \\ &= Pr(Y^1 \leq y^1|Y^0 \leq y^0, D=1) \times Pr(Y^0 \leq y^0|D=1). \end{aligned}$$

Assumption **A1** implies

$$Pr(Y^1 \leq y^1|Y^0 \leq y^0, D=1) = Pr(Y^1 \leq y^1|Y^0 \leq y^0, D=0).$$

Combining these results and again using the definition of a conditional probability distribu-

tion yields

$$\begin{aligned}
F_{Y^1, Y^0|D=1}^{CF}(y^1, y^0) &= Pr(Y^1 \leq y^1 | Y^0 \leq y^0, D=0) \times Pr(Y^0 \leq y^0 | D=1) \\
&= Pr(Y^1 \leq y^1 | Y^0 \leq y^0, D=0) \times Pr(Y^0 \leq y^0 | D=0) \times \frac{Pr(Y^0 \leq y^0 | D=1)}{Pr(Y^0 \leq y^0 | D=0)} \\
&= Pr(Y^1 \leq y^1, Y^0 \leq y^0 | D=0) \times \frac{Pr(Y^0 \leq y^0 | D=1)}{Pr(Y^0 \leq y^0 | D=0)} \\
&= F_{Y^1, Y^0|D=0}(y^1, y^0) \times \frac{F_{Y^0|D=1}(y^0)}{F_{Y^0|D=0}(y^0)}.
\end{aligned}$$

Assumption **A2** ensures that there are no values of y^0 for which the denominator in the final line is zero and the numerator is positive. Hence, $F_{Y^1, Y^0|D=1}^{CF}(y^1, y^0)$ is a well defined joint probability distribution. \square

The counterfactual joint distribution for the treatment group is constructed by rescaling the observed joint distribution in the control group. The scaling factor accounts for differences in the baseline distribution of outcomes between the two groups. If, for example, the treatment group includes more individuals with high Y^0 , the scaling factor assigns more weight to control individuals with high Y^0 than to those with low Y^0 .

Note that this result identifies the counterfactual joint distribution for a discrete or continuous outcome. Some mobility measures are only defined with additional assumptions on the outcome distribution. For example, transition matrices that report the probability p_{ij} that an individual in quintile i of pre-treatment income moves to quintile j of post-treatment income are only defined for strictly continuous outcomes. Identifying MTEs for certain mobility measures may therefore require additional assumptions. Theorem 1 therefore identifies MTEs on any mobility measure that is itself well-defined for a given type of data.

What are the substantive restrictions imposed by the identifying assumptions? Assumption **A1** is a special case of the unconfoundedness or selection on observed variables assumption. It is consistent with selection into groups based on levels of pre-treatment outcomes Y^0 or gains from treatment $\tilde{Y}^1 - Y^1$ but is violated by selection into groups based on time trends in outcomes $Y^1 - Y^0$. Consider the common treatment effects example of a job train-

ing program. MTEs can be identified if trainees have higher or lower average income than the control group or the effect of training is much smaller or larger than on the control group. Economic shocks between the pre- and post-treatment periods that affect trainees and the control group equally are not a problem. MTEs cannot be identified if the trainees' outcomes in the absence of training would have a different time trend to the control group. In particular, trainees opting into training due a transitory drop in their income would be a problem for identification (Ashenfelter, 1978).

Assumption **A2** is a common support condition. It ensures that for every possible pre-treatment outcome in the treatment group, there is a positive probability of observing the same outcome in the control group. If there is some value y^0 that may be observed in the treatment group but cannot be observed in the control group, then there is no counterfactual intertemporal transition that can be assigned to treated individuals with baseline outcome y^0 . This is a population assumption and so is not formally testable in any sample. However, it is possible to informally assess its plausibility by examining the marginal distributions of pre-treatment outcomes in the treatment and control groups. If the assumption fails, it is still possible to identify $F_{Y^1, Y^0|D=1}^{CF}(y^1, y^0)$ for those values of Y^0 observed in the control group.

3.2.1 A Toy Example

Applying the identification result to mobility measures based on transition matrices is straightforward. The transition matrix is a discretization of the conditional distribution of post-treatment outcomes given pre-treatment outcomes: $F_{Y^1|Y^0, D}(\cdot)$. So M^0 in equation (3.1) represents the conditional distribution

$$Pr(Y^1 \leq \bar{Y} | Y^0 \leq \bar{Y}, D = 0) = 1 - Pr(Y^1 \leq \bar{Y} | Y^0 > \bar{Y}, D = 0) = 2/3$$

and M^1 represents the conditional distribution

$$Pr(\tilde{Y}^1 \leq \bar{Y} | Y^0 \leq \bar{Y}, D = 1) = 1 - Pr(\tilde{Y}^1 \leq \bar{Y} | Y^0 > \bar{Y}, D = 1) = 1/3.$$

Under assumptions **A1** and **A2**, M^0 is equal to the conditional distribution of post-treatment outcomes that the treatment group would have obtained if treatment were not applied.

The transition matrix by itself is sufficient to construct some mobility measures. For example, the treatment effect on the probability of crossing the poverty line is

$$\Delta_1^{MTE} = \mu_1(M^1) - \mu_1(M^0) = \frac{2}{3} - \frac{1}{3} = \frac{1}{3}.$$

This calculation was possible only because the transition matrices are symmetric. Consider a non-symmetric transition matrix:

$$\begin{pmatrix} \frac{1}{3} & \frac{2}{3} \\ \frac{1}{3} & \frac{2}{3} \end{pmatrix}.$$

Here the probability of crossing the poverty line is $\pi \times \frac{2}{3} + (1 - \pi) \times \frac{1}{3}$, where π is the proportion of individuals below the poverty line in the pre-treatment period. This illustrates the general point that the transition matrix or conditional distribution of Y^1 given Y^0 is not sufficient for constructing many mobility measures. In general, the marginal distribution of Y^0 is also necessary. This marginal distribution also features in the rescaling factor in equation (3.4). In this case the rescaling factor would simply be $\frac{\pi_1}{\pi_0}$, the relative proportion of individuals below the poverty line in each group.

The discussion up to this point has treated \bar{Y} as a fixed parameter, such as the poverty line. However, many applications define transition matrices with respect to percentiles of the outcome distributions. Thus,

$$\begin{pmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \\ p_{41} & p_{42} & p_{43} & p_{44} \end{pmatrix}$$

might represent the set of conditional probabilities that an individual in quartile i of baseline income moves to quartile of j of subsequent income. The identification conditions above are unaffected by whether the cell boundaries are fixed or relative values. However, it is not *a priori* clear whether cell boundaries for the treatment group should be defined using the

treatment or control group outcomes. The former approach captures only within-group mobility, or the tendency of treatment to change the relative ranks of individuals in the treatment group. The latter approach captures both within-group rerankings and any level effects of the treatment. The cell boundaries could also be defined with respect to some broader population in a census or unrelated survey.

To concretize this issue, consider a treatment that raises the outcomes of all treated individuals by Δ . This will have zero mobility effect when cell boundaries are based on percentiles of the treatment group's outcomes but will have a positive mobility effect when cell boundaries are based on percentiles of the control group's outcomes. The former is arguably most appropriate if the researcher is interested in the effect of the treatment if it were applied to both treatment and control groups, whereas the latter is most appropriate if the treatment is only intended to be applied to part of the relevant population.

3.2.2 Covariates and Conditional Identification

A conditional analogue of theorem 1 applies when appropriate covariates X are available. Note that the identification result is entirely nonparametric and makes no assumptions about the functional form of the relationship between the covariates and the outcomes.

Theorem 2. *If*

B1 $Y^1 \perp D | Y^0, X$ *and*

B2 $\text{support}(Y^0 | D = 1, X) \subseteq \text{support}(Y^0 | D = 0, X)$

then the unobserved joint distribution of pre- and post-treatment outcomes in the treatment group $F_{Y^1, Y^0 | D=1, X}(\cdot, \cdot)$ equals the rescaled observed joint distribution of pre- and post-treatment outcomes in the control group

$$F_{Y^1, Y^0 | D=0, X}(y^1, y^0) \times \frac{\Pr(D = 1 | Y^0 \leq y^0, X)}{1 - \Pr(D = 1 | Y^0 \leq y^0, X)} \quad (3.5)$$

The rescaling factor now depends on differences between the treatment and control groups in the distribution of pre-treatment outcomes Y^0 and covariates X . This is the same function

employed in standard matching and reweighting methods for cross-sectional or longitudinal data. The conditional selection assumption **B1** requires that any differential time trend between the treatment and control groups' untreated outcomes be fully explained by differences in their covariates. I illustrate this point by returning to the job training example. Assume that 60% of the trainees are women and 40% are men, that the proportions are reversed in the control group and that men and women would have different trends in their outcomes in the absence of treatment. This would violate assumption **A1** but not assumption **B1**. Equation (3.5) shows that this pattern can be accounted for by estimating mobility treatment effects separately for men and women, and then taking a weighted average of these effects with weights 0.6 and 0.4 on the female and male effects respectively.

The conditional contained support assumption **B2** is stronger than its unconditional analogue, as it holds within each cell of the covariate grid, rather than across the full sample. Using the job training example, **B2** requires that there is contained support for the pre-treatment outcomes of each gender. If there exists a value y^0 that is found for men in the treatment group and women in both the treatment and control groups, **B2** will fail but **A2** will still hold.

As in any other method of program evaluation, only covariates that are not themselves determined by the treatment should be included in the reweighting exercise. This excludes many covariates that are measured after the treatment is implemented but may also exclude covariates that are measured between the treatment's announcement and implementation and are contaminated by anticipation effects.

3.2.3 Which research designs satisfy the identification conditions?

Here I consider how the identification assumptions **A1** and **B1** relate to common research designs used in applied microeconomic research.

Controlled randomization: The identifying assumptions **A1** and **B1** are satisfied if individuals are placed in the treatment or control group by random assignment or random assignment conditional on pre-treatment outcomes.

Difference-in-differences: These assumptions are strictly stronger than the "equal mean trends" identification condition in the linear difference-in-differences panel model de-

sign. These conditions are very similar to the identifying assumptions for the nonlinear difference-in-differences panel model proposed by Athey and Imbens (2006). Note that mobility treatment effects cannot be implemented using repeated cross-sections.

Regression discontinuity: The validity of the identifying assumptions depends on whether Y^0 is the running variable that determines assignment to treatment. If pre-treatment outcomes are used to determine assignment to treatment, then assumptions **A2** and **B2** will fail as the support of Y^0 will be entirely disjoint between the treatment and control groups. As an example, Matsudaira (2008) studies the effect on test scores of a remedial education program where entry into the program was determined by baseline test scores. If Y^0 is not the running variable and is not too strongly correlated with the running variable, assumptions **A2** and **B2** will not automatically fail. They should then hold locally under the same assumptions used to estimate average treatment effects in regression discontinuity designs (Lee and Lemieux, 2010).

Selection on observed variables: Assumption **A1** is satisfied whenever group assignment is uncorrelated with post-treatment outcomes conditional on pre-treatment outcomes $Y^1 \perp D|Y^0$. This is a special case of the selection on observed variables assumption $Y^1 \perp D|X$ where the only conditioning variable is Y^0 .

3.2.4 *Decomposition methods*

There may be interest in comparing mobility measures across two groups even when assumption **B1** fails and this comparison cannot be assigned a causal interpretation. In such cases, it is possible to decompose the difference into a “composition effect,” capturing differences in covariates and a “structure effect” capturing differences in the relationship between these covariates and mobility. In order to distinguish this from the program evaluation application discussed elsewhere in the paper, I denote the two groups in this application by $D \in \{A, B\}$. Here I treat group A as the reference group and apply equation (3.5) to calculate the counterfactual level of mobility in group B if it had the same distribution of baseline outcomes

and covariates as group A .

$$\underbrace{\mu(F_{Y^1, Y^0|D=A}(\cdot)) - \mu(F_{Y^1, Y^0|D=B}(\cdot))}_{\text{Total mobility difference}} = \underbrace{\mu(F_{Y^1, Y^0|D=A}(\cdot)) - \mu(F_{Y^1, Y^0|D=B}^{CF}(\cdot))}_{\text{Structure effect on mobility}} + \underbrace{\mu(F_{Y^1, Y^0|D=B}^{CF}(\cdot)) - \mu(F_{Y^1, Y^0|D=B}(\cdot))}_{\text{Composition effect on mobility}}$$

Note that $F_{Y^1, Y^0|D=B}^{CF}(\cdot)$ is only a well-defined distribution if assumption **B2** continues to hold, so this decomposition exercise should only be performed when there is at least some common support between the two groups. See Fortin, Lemieux, and Firpo (2011) for more discussion on this proviso and on interpreting the output of decomposition exercises.

3.2.5 More than two time periods

Multiple pre-treatment time periods can be used to test the identifying assumptions **A1** and **B1**. Specifically, these assumptions imply that in a stationary economic environment, the level of mobility between the two pre-treatment time periods, $T = -1$ and $T = 0$, should be identical in the treatment and control groups. Testing

$$H_0 : \quad \Delta^{\text{MTE}} = \mu(F_{Y^0, Y^{-1}|D=1}(\cdot, \cdot)) - \mu(F_{Y^0, Y^{-1}|D=0}(\cdot, \cdot)) = 0$$

establishes whether the identifying assumption holds in the pre-treatment periods. This test of the identifying assumption may yield erroneous conclusions if the economic environment is rapidly changing between periods -1 and 0 in ways that affect the two groups differently but is relatively stable between periods 0 and 1. With multiple time periods it may also be possible to select a pre-treatment time period $T = t$ such that a lagged identifying assumption $Y^1 \perp D|Y^t$ is more plausible than $Y^1 \perp D|Y^0$. Bell, Blundell, and van Reenen (1999) make a similar argument in the context of a difference-in-differences model.

3.2.6 More than two groups

Multiple treatment groups can each be analyzed relative to a single control group using equation (3.4) or (3.5). Each comparison yields a mobility treatment effect for the rele-

vant treatment group. If the different treatment groups received different treatments, the effects can be compared. If the different groups received different treatments, equality of the effects should be tested. Rejection of equality of the effects of the same treatment on different populations imply that the populations must differ on some observed or unobserved characteristic.

3.3 Estimation

Sufficient conditions for estimation of and inference on mobility measures have been established by prior authors, so I provide only a brief review of these techniques in this section.

Estimation of mobility treatment effects is a relatively straightforward exercise. All of the scalar summary measures of transition matrices that have been proposed as mobility measures can be written in the form $\mu(p_{ij}, \pi_i | i, j = 0, \dots, n)$ for some continuous function $\mu(\cdot)$. The p_{ij} and π_i terms are conditional and marginal probabilities that can be consistently estimated by sample analogues \hat{p}_{ij} and $\hat{\pi}_i$ under weak regularity conditions. The continuous mapping theorem then yields $\hat{\mu} \rightarrow_p \mu$.

Estimating dominance criteria that directly compare the joint distributions $F_{\tilde{Y}^1, Y^0|D=1}(\cdot)$ and $F_{Y^1, Y^0|D=1}^{CF}(\cdot)$ is slightly more complex. However, techniques for estimating distribution functions are now well established and, unlike most nonparametric estimators, they converge at parametric rates (Pagan and Ullah, 1999). For example, the counterfactual distribution in equation (3.4) can be consistently estimated by a product of observed empirical distribution functions

$$\frac{1}{N_0} \sum_{i=1}^{N_0} \mathbf{1}\{Y^1 \leq y_1, Y^0 \leq y^0\} \times \frac{\frac{1}{N_1} \sum_{i=1}^{N_1} \mathbf{1}\{Y^0 \leq y^0\}}{\frac{1}{N_0} \sum_{i=1}^{N_0} \mathbf{1}\{Y^0 \leq y^0\}}$$

where N_1 and N_0 are the sample sizes in the treatment and control groups respectively and $\mathbf{1}\{Z\}$ is an indicator function equal to one when condition Z is satisfied. In practice, most of the empirical literature appears to sidestep this challenge by using only scalar summary measures of mobility or evaluating the distribution functions at a finite and relatively small number of points.

The counterfactual distribution in equation (3.5) can be estimated in an analogous fash-

ion. The estimator can be written as:

$$\frac{1}{N_0} \sum_{i=1}^{N_0} \mathbf{1}\{Y^1 \leq y_1, Y^0 \leq y^0\} \times \frac{\hat{Pr}(D = 1|Y^0 \leq y^0, X)}{1 - \hat{Pr}(D = 1|Y^0 \leq y^0, X)}$$

The first term is simply an empirical distribution functions. The predicted probabilities $\hat{Pr}(D = 1|Y^0 \leq y^0, X)$ can be estimated by semiparametric logistic model.

A number of papers have established analytical variance formulae for mobility measures (Formby, Smith, and Zheng, 2004; Schluter, 1998). These can be applied directly to mobility treatment effects, although most of these approaches assume strictly continuous, independently distributed data. Biewen (2002) establishes the validity of the bootstrap for a number of scalar mobility measures under weak regularity conditions: a finite second moment of the outcomes and smoothness of the mobility measures with respect to the first moment of the outcomes. His approach allows for dependencies in the data and cluster-level resampling may be used to account for such dependence. Davidson and Flachaire (2007) argue that bootstrap-based tests may be very sensitive to outliers, in the sense that the bootstrap distribution of the test statistic depends heavily on the proportion of resamples in which these outliers are included. To illustrate why this may be a problem, note that estimates of equation (3.4) may be poorly behaved when $\frac{1}{N_0} \sum_{i=1}^{N_0} \mathbf{1}\{Y^0 \leq y^0\} \approx 0$.

3.4 Application

I conclude by presenting a brief application of mobility treatment effects to the evaluation of an education intervention in western Kenya. Duflo, Dupas, and Kremer (2011) study a group of 121 primary schools with two grade one classes. Schools are randomly assigned to either *divide students into high- and low-track classes* using their scores on a pre-treatment standardized test or *randomly divide students between two classes*. The original paper shows that after one year of treatment, students in tracking schools scored on average 0.14 standard deviations higher on standardized tests than students in random assignment schools. The paper also reports positive treatment effects for students in each quartile of baseline test scores, though the effects are larger for students in higher quartiles. The authors argue that this provides evidence that tracking benefits both low- and high-achieving students, with the

former group benefiting from more effective instruction in more homogeneous classrooms.

I argue that mobility treatment effects provide a useful complement to the evaluation strategies employed in this paper and permit a more nuanced evaluation of the experiment. The direction of the mobility treatment effects are not *a priori* clear in this setting. If peer effects are an important determinant of test scores, below-median students who are assigned to “low track” classrooms and hence to weaker peers may have a lower probability of catching up to “high track” students above the median than in random assignment schools. However, if instructors focus their attention on the highest scoring students in their class, then students marginally below the median may have a higher probability of catch-up in tracking schools than in random assignment schools.

I use a sample of 5170 students for whom both baseline and endline test scores are available. This sample includes 60 tracking schools, which I consider to be the treatment group, and 58 random assignment schools, which I consider to be the control group. I begin by constructing four-by-four transition matrices that measure the probability that students in quartile i of the baseline test score distribution reach quartile j of the endline test score distribution. I define the quartile boundaries using test scores in the control group. This yields the marginal distributions

$$\pi^1 = \begin{pmatrix} .27 & .25 & .24 & .24 \end{pmatrix}' \quad \pi^0 = \begin{pmatrix} .25 & .25 & .25 & .25 \end{pmatrix}'$$

for the treatment and control groups respectively. Note that the groups are not identically sized in the treatment group because the reference quantiles are defined using the control group and the distributions of baseline test scores are similar but not identical. The transition matrices for the two groups are given by

$$P^1 = \begin{pmatrix} .42 & .31 & .17 & .1 \\ .19 & .29 & .29 & .23 \\ .11 & .25 & .31 & .33 \\ .04 & .14 & .26 & .56 \end{pmatrix} \quad P^0 = \begin{pmatrix} .52 & .25 & .15 & .08 \\ .28 & .33 & .24 & .15 \\ .14 & .23 & .31 & .32 \\ .06 & .17 & .29 & .49 \end{pmatrix}.$$

A visual inspection of the two matrices suggests that mobility is slightly higher in the tracking

schools, particularly for the lower quartiles. This intuition can be formalized by computing two summary measures: the average probability of changing quartiles, defined as M_1 above, and the average number of quartile divisions crossed, defined as $M_2 = \sum_{i=1}^4 \sum_{j=1}^4 \pi_i p_{ij} |i - j|$. The results of this calculation are shown below, with cluster bootstrap standard errors in parentheses.

| | Treatment | Control | Difference |
|-------------------------------------|-----------|---------|---------------|
| Probability of changing quartile | .607 | .589 | .018** (.009) |
| Average number of quartiles changed | .831 | .813 | .018* (.011) |

This analysis confirms that tracking causes a marginally higher level of mobility than random assignment to classrooms. The effects are small, with the probability of changing quartiles rising by less than 2 percentage points, but precisely estimated. These estimates are based on quartiles of the control group's test scores and so combine rank changes within the treatment group and rank changes relative to the control group. I can also estimate the treatment group's transition matrix using quartiles of the treatment group's own distribution and so isolate the pure reranking effect of the treatment.

| | Treatment | Control | Difference |
|-------------------------------------|-----------|---------|-------------|
| Probability of changing quartile | .598 | .589 | .009 (.010) |
| Average number of quartiles changed | .826 | .813 | .012 (.011) |

The pure reranking effect within the treatment group accounts for approximately half of the total mobility treatment effect. The remainder is driven by the level change in the treatment group's outcomes relative to the control group.

I also implement a more direct test of the mobility hypotheses outlined above, using averaged within-school mobility. Specifically, I restrict the sample to students in the second and third quartiles of baseline test scores within each school, calculate the probability that a student below the school-specific median in the baseline will move above the median in the follow-up and *vice versa*, and test whether this probability differs between tracking and random assignment schools. This tests whether tracking increases or decreases the probability of reranking for students in the neighborhood of the cutoff separating low and

high track classrooms. The relevant transition matrices are:

$$P^1 = \begin{pmatrix} .57 & .43 \\ .43 & .57 \end{pmatrix} \quad P^0 = \begin{pmatrix} .61 & .39 \\ .39 & .61 \end{pmatrix}$$

so the probability of switching from above to below the median or *vice versa* is 4 percentage points higher in the control than the treatment group. This effect is marginally significant and suggests that tracking increased mobility by approximately seven percent of its baseline level amongst students in the neighborhood of the threshold separating high and low track classrooms.³

This result is consistent with the hypothesis that teachers focus their effort on students near the top of their classroom, which benefits students who marginally miss assignment to high track classrooms. Duflo, Dupas, and Kremer (2011) find that students just below and just above the cutoffs in tracking schools experience no difference in *levels* of their endline test scores. They interpret this as evidence that the effects on low track students of tailored instruction and weaker peers are approximately equal and offset each other. My results suggest that the former effect is more important when reranking is the outcome of interest and emphasize the potentially important role of mobility treatment effects as an evaluation criterion.

3.5 Conclusion

The measurement and analysis of intertemporal economic mobility is an important subject in both academic research and policy analysis. Intergenerational economic mobility has featured prominently in the practice and study of politics. Economists and other social scientists have studied how educational attainment, income, and occupation changes through time and between generations. Developing satisfactory measures of economic mobility has been a central theme of this research agenda.

This paper demonstrates how to interpret intertemporal economic mobility measures in a treatment effects or decomposition framework. I derive a simple and intuitive set of

³Restricting the sample to the fifth and six deciles yields a larger treatment effect of 7 percentage points or 14 percent of the baseline level of mobility, though this is very imprecisely estimated.

sufficient conditions for identifying mobility treatment effects, which I define as the causal effect of some treatment on mobility. I show how these conditions relate to common research designs in the applied microeconomics literature. The existing econometrics literature has already established sufficient conditions for estimation of mobility measures, which permits straightforward estimation of mobility treatment effects. Appropriate bootstrap procedures are valid for inference on most mobility treatment effects.

This framework will be useful in three settings. First, mobility measures provide a useful complement to existing treatment effects estimands, whether used in program evaluation or in more general economic analysis. Second, I clarify how to interpret differences in economic mobility across substantially different groups and link this interpretation into the decomposition literature. Third, mobility measures will be useful in the analysis of non-cardinal data such as norm-referenced standardized test scores.

Bibliography

- ABADIE, A. (2005): “Semiparametric Difference-in-difference Estimators,” *Review of Economic Studies*, 72, 1–19.
- AL-SAMARRAI, S., AND H. ZAMAN (2000): “Abolishing School Fees in Malawi: The Impact on Education Access and Equity,” *Education Economics*, 15(3), 359–375.
- AMMERMUELLER, A., AND J.-S. PISCHKE (2009): “Peer Effects in European Primary Schools: Evidence from PIRLS,” *Journal of Labor Economics*, 27(3), 315–348.
- ANGELUCCI, M., AND G. DI GIORGI (2009): “Indirect Effects of an Aid Program: How do Cash Injections Affect Ineligibles’ Consumption?,” *American Economic Review*, 99(1), 486–508.
- ANGRIST, J., AND K. LANG (2004): “Does School Integration Generate Peer Effects? Evidence from Boston’s Metco Program,” *American Economic Review*, 94(5), 1613–1634.
- ARNOTT, R. (1987): “Peer Group Effects and Educational Attainment,” *Journal of Public Economics*, 32, 287–305.
- ASHENFELTER, O. (1978): “Estimating the Effect of Training Programs on Earnings,” *Review of Economics and Statistics*, 60(1), 45–57.
- ATHEY, S., AND G. IMBENS (2006): “Identification and Inference in Nonlinear Difference-in-differences Models,” *Econometrica*, 74(2), 431–497.
- ATKINSON, A. (1983): “The Measurement of Economic Mobility,” in *Social Justice and Public Policy*, ed. by A. Atkinson. MIT Press, Cambridge, MA.
- BANERJEE, A., S. GALIANI, J. LEVINSOHN, Z. McLAREN, AND I. WOOLARD (2008): “Why Has Unemployment Risen in the New South Africa?,” *The Economics of Transition*, 16(4), 715–740.
- BARRERA-OSORIO, F., L. LINDEN, AND M. URQUIOLA (2007): “The Effects of User Fee Reductions on Enrollment: Evidence from a Quasi-experiment,” Mimeo, Columbia University.
- BARTHOLOMEW, D. (1982): *Stochastic Models for Social Processes*. Wiley, London, 3 edn.
- BECKER, G., AND K. MURPHY (2000): *Social economics: Market behavior in a social environment*. Harvard University Press, Cambridge, MA.
- BEHRMAN, J. R., A. D. FOSTER, M. R. ROSENZWEIG, AND P. VASHISHTHA (1999): “Women’s Schooling, Home Teaching, and Economic Growth,” *Journal of Political Economy*, 107(4), 682–715.

- BELL, B., R. BLUNDELL, AND J. VAN REENEN (1999): “Getting the Unemployed Back to Work: The Role of Targeted Wage Subsidies,” *International Tax and Public Finance*, 6(3), 339–360.
- BENABOU, R. (1996): “Equity and Efficiency in Human Capital Investment: The Local Connection,” *Review of Economic Studies*, 63(2), 237–264.
- BERTRAND, M., E. DUFLO, AND S. MULLAINATHAN (2004): “How Much Should We Trust Difference-in-differences Estimates?,” *Quarterly Journal of Economics*, 119(1), 249–275.
- BERTRAND, M., R. HANNA, AND S. MULLAINATHAN (2010): “Affirmative Action in Education: Evidence from Engineering College Admissions in India,” *Journal of Public Economics*, 94(1/2), 16–29.
- BESLEY, T., AND A. CASE (1993): “Modeling Technology Adoption in Developing Countries,” *American Economic Review*, 83(2), 396–402.
- BETTS, J. (2011): “The Economics of Tracking in Education,” in *Handbook of the Economics of Education Volume 3*, ed. by E. Hanushek, S. Machin, and L. Woessmann, pp. 341–381. Elsevier.
- BHATTACHARYA, D. (2009): “Inferring Optimal Peer Assignment from Experimental Data,” *Journal of the American Statistical Association*, 104(486), 486–500.
- BIEWEN, M. (2002): “Bootstrap Inference for Inequality, Mobility and Poverty Measurement,” *Journal of Econometrics*, 108, 317–342.
- BLUME, L., W. BROCK, S. DURLAUF, AND Y. IOANNIDES (2011): “Identification of Social Interactions,” in *Handbook of Social Economics Volume 1B*, ed. by J. Benhabib, A. Bisin, and M. Jackson, pp. 853–964. Elsevier.
- BOISJOLY, J., G. DUNCAN, M. KREMER, D. LEVY, AND J. ECCLES (2006): “Empathy or Antipathy: The Impact of Diversity,” *American Economic Review*, 95(5), 1890–1905.
- BORKUM, E. (2011): “Can Eliminating School Fees in Poor Districts Boost Enrollment? Evidence from South Africa,” *Economic Development and Cultural Change*, Forthcoming.
- BRANSON, N., D. LAM, AND L. ZUZE (2012): “Education: Analysis of the NIDS Wave 1 and 2 Datasets,” Discussion Paper 81, Southern Africa Labour and Development Research Unit.
- BROCK, W., AND S. DURLAUF (2011): “Interactions-based Models,” in *Handbook of Econometrics Volume 5*, ed. by J. Heckman, and E. Leamer, pp. 3297–3380. Elsevier.
- BURMAN, P., E. CHOW, AND D. NOLAN (1994): “A Cross-Validatory Method for Dependent Data,” *Biometrika*, 81(2), 351–358.
- CAMERON, C., D. MILLER, AND J. GELBACH (2008): “Bootstrap-based Improvements for Inference with Clustered Errors,” *Review of Economics and Statistics*, 90(3), 414–427.
- CARD, D. (1999): “The Causal Effect of Education on Earnings,” in *Handbook of Labor Economics Volume 3A*, ed. by D. Card, and O. Ashenfelter, pp. 1801–1863. Elsevier.
- CARRELL, S., R. FULLERTON, AND J. WEST (2009): “Does Your Cohort Matter? Measuring Peer Effects in College Achievement,” *Journal of Labor Economics*, 27(3), 439–464.

- CARRELL, S., B. SACERDOTE, AND J. WEST (2012): “From Natural Variation to Optimal Policy? An Unsuccessful Experiment in Using Peer Effects Estimates to Improve Student Outcomes,” Working paper.
- CENTRE FOR DEVELOPMENT ENTERPRISE (2010): “Hidden Assets: South Africa’s Low-Fee Private Schools,” .
- COOLEY, J. (2010): “Can Achievement Peer Effect Estimates Inform Policy? A View from Inside the Black Box,” Working paper.
- CULLEN, J., M. LONG, AND R. REBACK (2013): “Jockeying for Position: Strategic High School Choice under Texas’ Top Ten Percent Plan,” *Journal of Public Economics*, 97, 32–48.
- DAVIDSON, R., AND E. FLACHAIRE (2007): “Asymptotic and Bootstrap Inference for Inequality and Poverty Measures,” *Journal of Econometrics*, 141(1), 141–166.
- DEININGER, K. (2003): “Does the Cost of Schooling Affect Enrollment by the Poor? Universal Primary Education in Uganda,” *Economics of Education Review*, 22, 291–305.
- DEPARTMENT OF EDUCATION (2009): *Trends in Education Macro Indicators Report*. Government Printers, Pretoria, ZA.
- (2011): *Trends in Education Macro Indicators Report*. Government Printers, Pretoria, ZA.
- DI GIORGI, G., M. PELLIZZARI, AND S. REDAELLI (2010): “Identification of Social Interactions through Partially Overlapping Peer Groups,” *American Economic Journal: Applied Economics*, 2(2), 241–275.
- DINARDO, J., N. FORTIN, AND T. LEMIUEX (1996): “Labor Market Institutions and the Distribution of Wages, 1973 - 1992: A Semiparametric Approach,” *Econometrica*, 64(5), 1001–1044.
- DING, W., AND S. LEHRER (2007): “Do Peers Affect Student Achievement in China’s Secondary Schools?,” *Review of Economics and Statistics*, 89(2), 300–312.
- DUFLO, E. (2001): “Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment,” *American Economic Review*, 91(4), 795–813.
- DUFLO, E., P. DUPAS, AND M. KREMER (2011): “Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya,” *American Economic Review*, 101(5), 1739–1774.
- DUPAS, P. (2013): “Short-Run Subsidies and Long-Run Adoption of New Health Products: Evidence from a Field Experiment,” Mimeo.
- DYNARSKI, S. (2003): “Does Aid Matter? Measuring the Effect of Student Aid on College Attendance and Completion,” *American Economic Review*, 93(1), 279–288.
- DYNARSKI, S., J. GRUBER, AND D. LI (2009): “Cheaper by the Dozen: Using Sibling Discounts at Catholic Schools to Estimate the Price Elasticity of Private School Attendance,” Discussion Paper 15461, National Bureau of Economic Research.

- EDMONDS, E. (2006): "Child Labor and Schooling Responses to Anticipated Income in South Africa," *Journal of Development Economics*, 81(2), 386–414.
- EPPLE, D., AND R. ROMANO (1998): "Competition Between Private and Public Schools, Vouchers and Peer-Group Effects," *American Economic Review*, 88(1), 33–62.
- (2011): "Peer Effects in Education: A Survey of the Theory and Evidence," in *Handbook of Social Economics Volume 1B*, ed. by J. Benhabib, A. Bisin, and M. Jackson, pp. 1053–1163. Elsevier.
- EVANS, D., M. KREMER, AND M. NGATIA (2009): "The Impact of Distributing School Uniforms on Children's Education in Kenya," Mimeo.
- FAFCHAMPS, M., AND B. MINTEN (2007): "Public service provision, user fees and political turmoil," 16(3), 485–518.
- FAN, J., AND I. GIJBELS (1996): *Local Polynomial Modelling and Its Applications*. Chapman & Hall.
- FEDDERKE, J., J. LUIZ, AND R. DE KADT (2000): "Uneducating South Africa: The Failure to Address the 1910–1993 Legacy," *International Review of Education*, 46(3/4), 257–281.
- FIELDS, G., AND E. OK (1996): "The Meaning and Measurement of Income Mobility," *Journal of Economic Theory*, 71, 349–377.
- FILMER, D., AND N. SCHADY (2008): "Getting Girls into School: Evidence from a Scholarship Program in Cambodia," *Economic Development and Cultural Change*, 56(3), 581–617.
- FIRPO, S. (2007): "Efficient Semiparametric Estimation of Quantile Treatment Effects," *Econometrica*, 75(1), 259–276.
- (2010): "Identification and Estimation of Distributional Impacts of Interventions Using Changes in Inequality Measures," IZA Discussion Paper 4841.
- FIZBEIN, A., AND N. SCHADY (2009): *Conditional Cash Transfers: Reducing Current and Future Poverty*. World Bank.
- FORMBY, J., J. SMITH, AND B. ZHENG (2004): "Mobility Measurement, Transition Matrices and Statistical Inference," *Journal of Econometrics*, 120, 181–205.
- FORTIN, N., T. LEMIUEX, AND S. FIRPO (2011): "Decomposition Methods in Economics," in *Handbook of Labor Economics Volume 4A*, ed. by O. Ashenfelter, and D. Card. North-Holland.
- FOSTER, G. (2006): "It's Not Your Peers and it's Not Your Friends: Some Progress Toward Understanding the Educational Peer Effect Mechanism," *Journal of Public Economics*, 90, 1455–1475.
- FRISANCHO, V., AND K. KRISHNA (2011): "Affirmative Action in Higher Education in India: Targeting, Catch Up, and Mismatch," Working paper.
- GALOR, O., AND D. TSIDDON (1997): "Technological Progress, Mobility, and Economic Growth," *American Economic Review*, 87(3), 363–382.

- GARLICK, R. (2012): “Mobility Treatment Effects: Identification, Estimation and Application,” Working paper.
- GLEWWE, P. (1999): “Why Does Mother’s Schooling Raise Child Health in Developing Countries? Evidence from Morocco,” *Journal of Human Resources*, 34(1), 124–159.
- GLEWWE, P., AND A. L. KASSOUF (2011): “The Impact of the Bolsa Escola/Familia Conditional Cash Transfer Program on Enrollment, Drop Out Rates and Grade Promotion in Brazil,” University of Minnesota.
- GLEWWE, P., AND M. KREMER (2006): “Schools, Teachers and Education Outcomes in Developing Countries,” in *Handbook of the Economics of Education, Volume 2*, ed. by E. Hanushek, and F. Welch, pp. 945–1017. North-Holland.
- GRAHAM, B. (2011): “Econometric Methods for the Analysis of Assignment Problems in the Presence of Complementarity and Social Spillovers,” in *Handbook of Social Economics Volume 1B*, ed. by J. Benhabib, A. Bisin, and M. Jackson, pp. 965–1052. Elsevier.
- GRAHAM, B., G. IMBENS, AND G. RIDDER (2011): “Measuring the Average Outcome and Inequality Effects of Segregation in the Presence of Social Spillovers,” Mimeo.
- GURRYAN, J., B. JACOB, E. KLOPFER, AND J. GROFF (2008): “Using Technology to Explore Social Networks and Mechanisms Underlying Peer Effects in Classrooms,” *Developmental Psychology*, 44(2), 355–364.
- HAHN, J. (1998): “On the Role of Propensity Score in Efficient Semiparametric Estimation of Average Treatment Effects,” *Econometrica*, 66, 315–332.
- HANES, S. (2006): “Easing the Burden of School Fees in Africa,” Christian Science Monitor.
- HANUSHEK, E., J. KAIN, J. MARKMAN, AND S. RIVKIN (2008): “Does Peer Ability Affect Student Achievement?,” *Journal of Applied Econometrics*, 18(5), 527–544.
- HANUSHEK, E., AND D. KIMKO (2000): “Schooling, Labor Force Quality, and the Growth of Nations,” *American Economic Review*, 90(5), 1184–1208.
- HECKMAN, J., AND J. HOTZ (1989): “Choosing Among Alternative Nonexperimental Methods for Estimating the Impact of Social Programs: The Case of Manpower Training,” *Journal of the American Statistical Association*, 84(408), 862–880.
- HECKMAN, J., L. LOCHNER, AND P. TODD (2006): “Earnings functions, rates of return and treatment effects: The Mincer equation and beyond,” in *Handbook of the Economics of Education Volume 2*, ed. by E. Hanushek, and F. Welch. North-Holland.
- HECKMAN, J., AND R. ROBB (1985): “Alternative Methods for Estimating the Impact of Interventions,” in *Longitudinal Analysis of Labor Market Data*, ed. by J. Heckman, and B. Singer. Cambridge University Press.
- HECKMAN, J., J. SMITH, AND N. CLEMENTS (1997): “Making the Most out of Programme Evaluations and Social Experiments: Accounting for Heterogeneity in Programme Impacts,” *Review of Economic Studies*, 64(4), 487–535.

- HIDALGO, D., M. ONOFA, H. OOSTERBEEK, AND J. PONCE (2013): “Can Provision of Free School Uniforms Harm Attendance? Evidence from Ecuador,” *Journal of Development Economics*, 103(3), 43–51.
- HIRANO, K., G. IMBENS, AND G. RIDDER (2003): “Efficient Estimation of Average Treatment Effects Using the Propensity Score,” *Econometrica*, 71(4), 1161–1189.
- HOROWITZ, J. (2001): “The Bootstrap,” in *The Handbook of Econometrics Volume 5*, ed. by J. Heckman, and E. Leamer, pp. 3159–3228. Elsevier.
- HOXBY, C. (2000): “Peer Effects in the Classroom: Learning from Gender and Race Variation,” Working paper 7867, National Bureau of Economic Research.
- HOXBY, C., AND G. WEINGARTH (2006): “Taking Race out of the Equation: School Reassignment and the Structure of Peer Effects,” Working paper.
- HSIEH, C.-T., AND M. URQUIOLA (2006): “The Effects of Generalized School Choice on Achievement and Stratification: Evidence from Chile’s Voucher Program,” *Journal of Public Economics*, 90(8-9), 1477–1503.
- IMBENS, G., AND K. KALYANARAMAN (2009): “Optimal Bandwidth Choice for the Regression Discontinuity Estimator,” Discussion Paper 14726, National Bureau of Economic Research.
- IMBENS, G., AND J. WOOLDRIDGE (2009): “Recent Developments in the Econometrics of Program Evaluation,” *Journal of Economic Literature*, 47(1), 5–86.
- IMBERMAN, S., A. KUGLER, AND B. SACERDOTE (2012): “Katrina’s Children: Evidence on the Structure of Peer Effects from Hurricane Evacuees,” *American Economic Review*, 102(5), 2048–2082.
- KANE, T. (1994): “College Entry by Blacks since 1970: The Role of College Costs, Family Background, and the Returns to Education,” *Journal of Political Economy*, 102(5), 878–911.
- KANG, C. (2007): “Classroom Peer Effects and Academic Achievement: Quasi-Experimental Evidence from South Korea,” *Journal of Urban Economics*, 61, 458–495.
- KLING, J., D. LIEBMAN, AND L. KATZ (2007): “Experimental Analysis of Neighborhood Effects,” *Econometrica*, 75(1), 83–119.
- KREMER, M., E. MIGUEL, AND R. THORNTON (2009): “Incentives to Learn,” *Review of Economics and Statistics*, 91(3), 437–456.
- LAM, D., C. ARDINGTON, N. BRANSON, K. GOOSTREY, AND M. LEIBBRANDT (2010): “Credit Constraints and the Racial Gap in Post-secondary Education in South Africa,” University of Michigan.
- LAM, D., C. ARDINGTON, AND M. LEIBBRANDT (2010): “Schooling as a Lottery: Racial Differences in School Advancement in Urban South Africa,” *Journal of Development Economics*, 95(2), 121–136.
- LAVY, V., AND A. SCHLOSSER (2011): “Mechanisms and Impacts of Gender Peer Effects at School,” *American Economic Journal: Applied Economics*, 3(2), 1–33.

- LEE, D., AND T. LEMIEUX (2010): “Regression Discontinuity Designs in Economics,” *Journal of Economic Literature*, 48(2), 281–355.
- LUCAS, A., AND I. MBITI (2009): “The Effect of Free Primary Education on Student Participation, Stratification and Achievement: Evidence from Kenya,” Mimeo, Wellesley University.
- MANKIW, G., D. ROMER, AND D. WEIL (1992): “A Contribution to the Empirics of Economic Growth,” *Quarterly Journal of Economics*, 107(2), 407–437.
- MANSKI, C. (1993): “Identification of Endogenous Social Effects: The Reflection Problem,” *Review of Economics and Statistics*, 60(3), 531–542.
- MARMAROS, D., AND B. SACERDOTE (2002): “Peer and Social Networks in Job Search,” *European Economic Review*, 46(4-5), 870–879.
- (2006): “How do Friendships Form?,” *Quarterly Journal of Economics*, 121, 79–119.
- MATSUDAIRA, J. (2008): “Mandatory Summer School and Student Achievement,” *Journal of Econometrics*, 142(2), 829–850.
- MCCRARY, J. (2008): “Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test,” *Journal of Econometrics*, 142(2), 698–714.
- NEAL, D. (2002): “How Vouchers Could Change the Market for Education,” *Journal of Economic Perspectives*, 16(4), 25–44.
- OZIER, O. (2011): “The Impact of Secondary Schooling in Kenya: A Regression Discontinuity Analysis,” Mimeo.
- PAGAN, A., AND A. ULLAH (1999): *Nonparametric Econometrics*. Cambridge University Press, Cambridge, UK.
- PAMPALLIS, J. (2008): *School Fees*. Centre for Education Policy Development, Johannesburg, ZA.
- POP-ELECHES, C., AND M. URQUIOLA (2012): “Going to a Better School: Effects and Behavioral Responses,” *American Economic Review*, forthcoming.
- PRAIS, S. (1955): “Measuring Social Mobility,” *Journal of the Royal Statistical Society Series A*, 118, 56–66.
- ROBINS, J., AND A. ROTNITZKY (1995): “Semiparametric Efficiency in Multivariate Regression Models with Missing Data,” *Journal of the American Statistical Association*, 90(429), 122–129.
- ROBINSON, P. (1988): “Root-n Consistent Semiparametric Regression,” *Econometrica*, 56(4), 931–954.
- SACERDOTE, B. (2001): “Peer Effects with Random Assignment: Results for Dartmouth Roommates,” *Quarterly Journal of Economics*, 116(2), 681–704.
- (2011): “Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far?,” in *Handbook of the Economics of Education Volume 3*, ed. by E. Hanushek, S. Machin, and L. Woessmann, pp. 249–277. Elsevier.

- SCHLUTER, C. (1998): "Statistical Inference with Mobility Indices," *Economics Letters*, 59, 157–162.
- SCHULTZ, P. (2004): "School Subsidies for the Poor: Evaluating the Mexican PROGRESA Poverty Program," 74(1), 199–250.
- SEEKINGS, J., AND N. NATTRASS (2005): *Class, Race and Inequality in South Africa*. Yale University Press, New Haven, CT.
- SEFTOR, N., AND S. TURNER (2002): "Back to School: Federal Student Aid Policy and Adult College Enrollment," *Journal of Human Resources*, 37(2), 336–352.
- SHORROCKS, A. (1978a): "Income Inequality and Income Mobility," *Journal of Economic Theory*, 19, 376–393.
- (1978b): "The Measurement of Mobility," *Econometrica*, 46, 10143–1024.
- SOLON, G. (1999): "Intergenerational Mobility in the Labor Market," in *Handbook of Labor Economics Volume 3*, ed. by O. Ashenfelter, and D. Card, pp. 1761–1800. Elsevier.
- (2002): "Cross-Country Differences in Intergenerational Earnings Mobility," *Journal of Economic Perspectives*, 16(3), 59–66.
- SOMMERS, P., AND J. CONLISK (1979): "Eigenvalue Immobility Measures for Markov Chains," *Journal of Mathematical Sociology*, 6, 253–276.
- STINEBRICKNER, T., AND R. STINEBRICKNER (2006): "What Can Be Learned About Peer Effects Using College Roommates? Evidence from New Survey Data and Students from Disadvantaged Backgrounds," *Journal of Public Economics*, 90(8/9), 1435–1454.
- (2008): "The Causal Effect of Studying on Academic Performance," *B.E. Journal of Economic Analysis and Policy*, 8(1), 1868–1868.
- TERREBLANCHE, S. (2002): *A History of Inequality in South Africa*. University of Natal Press, Pietermaritzburg.
- TODD, P., AND K. WOLPIN (2006): "Assessing the Impact of a School Subsidy Program in Mexico: Using a Social Experiment to Validate a Dynamic Behavioral Model of Child Schooling and Fertility," *American Economic Review*, 96(5), 1384–1417.
- VAN DEN BERG, S., AND M. LOUW (2007): "Lessons Learnt From SACMEQII: South African Student Performance in Regional Context," Discussion Paper 16, Stellenbosch University.
- VIGDOR, J., AND T. NECHYBA (2007): "Peer Effects in North Carolina Public Schools," in *Schools and the Equal Opportunity Problem*, ed. by L. Woessmann, and P. Peterson, pp. 73 – 102. MIT Press.
- WILDEMAN, A. (2008): "Reviewing Eight Years of the Implementation of School Funding Norms," Idasa Economic Government Programme Research Paper.
- YATCHEW, P. (1997): "An Elementary Estimator of the Partial Linear Model," *Economics Letters*, 57, 135–143.

ZIMMERMAN, D. (2003): “Peer Effects in Academic Outcomes: Evidence From a Natural Experiment,” *Review of Economics and Statistics*, 85(1), 9–23.